DIGITAL VIDEO FOR TIME BASED ANALYSIS SYSTEMS

A THESIS

SUBMITTED IN PARTIAL FULFILMENT

OF THE REQUIREMENTS FOR THE DEGREE

OF

MASTER OF SCIENCE IN COMPUTER SCIENCE

IN THE

UNIVERSITY OF CANTERBURY

by

Lindsay Brian Weir

University of Canterbury

1994

Abstract

Research students within the Psychology Department at the University of Canterbury are involved in exploring emotional responses of human behaviour. Experiments of subjects are video taped and the tapes are subsequently analysed using pen and paper. This approach is time consuming and provides a relatively crude interface for analysis. In this thesis techniques to assist in the analysis of time dependent information are examined in general, although the emphasis is on human behaviour experiments. Digital video analysis methods are examined to evaluate their strengths and weaknesses in comparison to video tape methods.

A working prototype system, *Video Transcriptor,* has been developed on a Macintosh computer in order to evaluate how digital video can assist in analysing human behaviour. This prototype system uses the facilities of QuickTime, Apple's solution to handling time based digital video information. There is a lack of standards for controlling digital video information, so an analysis of various Human-Computer Interface metaphors has been explored. For transcription purposes, an adaptive note-taking facility has been implemented to assist in the analysis of human behaviour.

This thesis shows the benefits that digital video provides for the analysis and note-taking of human behaviours compared to video tape methods. The random access capabilities of digital video offer increased control of the video information, which provides faster note-taking and more accurate results compared to video tape based methods of analysis.

Acknowledgements

I would like to thank my supervisor Dr. Tim Bell for his valued assistance throughout this research. My gratitude also to Pari Kolahi from the Psychology Department who has been involved in the design and testing of my research. I would record the support of my fellow postgraduate students, Andy Cockburn for background information, Hugh Emberson for statistical information, Jonathan Woodford-Robinson for television broadcasting information, Marcus Finlay for helping in the design and testing of *Video Transcriptor* , Tim Carey who assisted in proof-reading, and David Bainbridge for late night entertainment.

Contents

Chapter 1

Introduction

Research students within the Psychology Department at the University of Canterbury are attempting to understand emotional states of subjects from their facial expressions, body movements, and voice. The method currently used by the research students involves recording the subject onto video tape for later analysis using pen and paper. This type of video tape analysis has discouraged many researchers from using video in their work. The focus of this thesis is to explore alternative methods of analysing video, or any other type of time based information, through digital video. It is hoped that this alternative method will provide faster, easier, and more powerful facilities to assist in analysing time based information.

In order to fully appreciate what time based analysis techniques are, we have to first define what time based information is. Time based information is anything that can change over a period of real time. This information can consist of almost anything imaginable. Results of a chemical experiment over a fixed time interval, financial results of a business over the last five years, plant growth in a densely populated forest, human interaction in everyday life, all contain information that changes over a period of time. The period of time for each of these may be entirely distinct: a chemical experiment may be finished within half a second while a plant's growth may take one hundred years.

Fast chemical reactions are difficult, if not impossible, to observe with the naked eye. Analysis is normally carried out by calculating the expected reaction's results and then checking the actual results achieved. Video taping the chemical reaction may assist in visualising the reaction that occurs. Measuring equipment may also be employed to detect various changes in the chemical reaction.

A business on the other hand may wish to examine its financial standing over the last five years to see how well it is performing. Static graphs are a common method used to show results of a single or multiple variables over the period of time.

Analysis of plant growth is a very slow process. This can take anywhere between a few days to many years, depending on which aspect of the plant's life cycle is being analysed. A researcher may take down notes of the plants progress since the last observation and take photographs to show the changes over the period of time.

Analysis of human interaction in everyday life can be done in many ways. Taking notes while observing the human interaction requires minimal resources but there is a high potential for losing important information. Other methods such as code based encoders, discussed in chapter 2, may be employed to assist in the speed of analysing the human interaction. A common method being employed, due to the relative low cost, is the use of video tape recorders to capture human interaction so that it can be analysed later.

These examples show that the analysis method chosen depends on the type of information to be examined and the resources available.

1.1 Motivation

The last half of the twentieth century has seen phenomenal developments in human communication. Television has achieved a virtually universal presence in developed countries, and the computer is fast becoming the next major presence. Bringing the two technologies together could have a significant impact on how we work and play.

A significant difficulty preventing the integration of these two technologies stems from the fact television and video signals are analog based signals while computer technology is digital based. These two technologies do not mix easily. Analog video signals have to be converted to digital form before the computer can use them.

Until recently the integration of these two distinct technologies was rare. Now that it is possible and relatively inexpensive it can be exploited in many new ways. Digital broadcasting of television programs and interactive television are examples that are beginning to exploit digital video technology [29, 44]. One of these ways not yet explored is using digital video to assist in analysing human behaviour, at an affordable cost.

A variety of different methods for analysing time based information are examined to find out their individual strengths and weaknesses. Some of the methods for analysing time based information include: viewing the information in real time while taking observational notes with pen and paper, keypad based encoders, video tape methods including pen and paper note-taking procedures, computer controlled video tape players with note-taking facilities, and other digital based systems. Each of the different time based analysis methods offer various advantages and disadvantages. Cost is significant as the facilities become more powerful and advanced.

Motion based digital video technology for the mass of users is relatively new. In the early 1990's hardware developers were offering video digitising cards that they hoped would dominate the multimedia market. However, since these products targeted different niches in the market place and offered different feature sets, there was much chaos and redundancy in the market place. Each of these proprietary schemes had its own Application Programming Interface (API) causing enormous problems for software developers who wanted to support more than one type of digitising card for their program. The software developers had to either write one version of the application for each different video digitising card or were required to support every card in the one application. The inefficiency of this situation had kept developers from introducing innovative products in the digital video market quickly. Users were confused by the number of existing proprietary standards that were available, the vast differences in the user interfaces of various products, and in the capabilities of the hardware.

The introduction of QuickTime in December 1991, [20], changed all of this on the Apple Macintosh by providing a standard interface for dynamic data types. The modular nature of QuickTime allows hardware and software developers to concentrate on making their hardware and software, secure in the knowledge that their products will work with whatever Macintosh computers are in the marketplace. Application writers can at last write code that does not depend on the underlying hardware and can take advantage of new hardware features as they come along.

With the availability of widely accepted standards for digital video, the features can be exploited in many areas. One of the areas to be exploited and evaluated is the use of digital video as a tool to assist in the analysis of human behaviour and any other time based information. Digital video has many things to offer. These features will be examined against traditional methods including analysis methods involving video tape players.

Limitations of motion based digital video will also be examined. Digital video information involves tremendous amounts of information to be processed. A single picture contains approximately one megabyte of information. Full motion video contains 25 frames per second[1]. This amounts to approximately 25 megabytes of information to be processed per second. Hard disk drives cannot transfer this amount of information per second. Based on these figures, thirty seconds of digital video would take approximately 750 megabytes of disk storage. The only conceivable method of handling this amount of information is through compression. Therefore compression and playback methods play an important role when dealing with motion video information, and these will be examined in further detail.

Since digital video is in its infancy as an analysis tool, several Human-Computer Interface issues need to be examined. Some guide-lines need to be drawn up as to peoples likes, dislikes, and assumptions about an interactive system designed to assist with analysing time based information such as human behaviour. These guidelines need to cover such areas as the control of the digital video information and the human interaction with various computer based note-taking facilities.

1.2 Psychology Experiments

The present research for some students within the Psychology Department focuses on the deception in emotional and non-verbal behaviour through the analysis of facial expressions.

Experiments are performed on undergraduate students and video taped for later analysis. The experiments performed required a large number of subjects in order to produce accurate results. For each subject analysed, additional discussions are held to find out their responses to certain stimuli. Current analysis techniques, using video tape with pen and paper for note-taking, take a considerable amount of time. Video tape has many limitations, including the time required for analysing its contents, that hinder the note-taking process.

The research, development, and testing of a digital video analysis system has been carried out in conjunction with the Psychology Department. The objective has been to find out what benefits digital video offer over video tape in the analysis of human behaviour.

---

[1] Phase Alternating Line (PAL) standard as used in New Zealand.

There is an overwhelming amount of data to be processed. From this wealth of information only small segments of video tape information are analysed. This analysis consists of the researcher sitting in front of a video player with pen and paper taking notes of the events that occur. Once each video segment has been analysed, a correlation of results must be obtained. Since the results are compared on paper and not from the original source material the accuracy of results may suffer due to the inability to check them against the primary source of information.

The objective of this thesis is to compare a digital video based analysis system with a typical video tape scenario currently used by researchers. Digital video technology can offer several advantages over analog based systems in the analysis of the information stored. These will be examined.

The experimental room used is relatively bare without distractions for the subject. The subject is shown a pre-recorded video tape, chosen specifically, due to the emotion stimulating content. As the subject is watching the contents of the video, a camera records their facial expressions. Behind the television the entire wall is covered with a set of curtains. Behind this set of curtains there is an observational room located behind one-way glass for the camera. In this observational room, a second video camera is positioned to record the full body movements of the subject. The curtains between the observational and experimental rooms are slightly pulled apart in the middle to reveal a small section of the one-way glass. This is done as inconspicuously as possible, so as not to alarm the subject.

The two video signals from the cameras are merged onto a single video tape together with timing information, accurate to one tenth of a second. In comparison, video tape players only have counters down to one second accuracy. Figure 1.1 shows an example of the video and timing information recorded. No picture has been supplied due to the confidential nature of the video tapes.

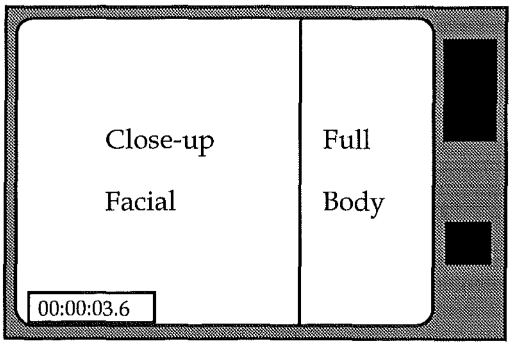


Figure 1.1. Video Tape Format Showing the Three Main Visual Areas.

After the experiments finish, the researcher goes through all of the tapes of the different subjects and selects a three minute section of video for each. Each subject typically has half an hour of material recorded during the experiments, most of which is irrelevant to the research. The three minute sections of video chosen for each subject are recorded onto a separate video tape from where the analysis takes place.

The quality of the video signal is impaired for five reasons: the inadequate lighting in the experimental room, old video camera technology, the use of one-way glass between the subject and one of the video cameras, the second generation copy of video being analysed, and the video tape media it is recorded on.

The experiments were carried out with only soft lighting which was necessary to set a relatively relaxed atmosphere for the subjects but consequently affected the quality of the video that was recorded. Bright lights can affect the subjects natural responses in the experiments.

The old video cameras also contributed to the poor video signal that had to be analysed. Improved picture signals are possible with new equipment, but it is a fact of life that researchers must work on a limited budget.

The one-way glass located between the subject and the second camera also contributed to reduced quality of the images recorded. Subsequently when the three minute section of the original tape is copied onto the new tape the quality of the video information is significantly reduced making it even more difficult to analyse.

The resulting video tape produced from the two video cameras is later examined by the researcher to extract information about behavioural characteristics exhibited by the subject. These results are compared between all the subject's that have been examined under the same conditions. From this analysis the researcher can find correlations between the contents of the tape shown to the subjects, with the emotions expressed by the subjects. The accuracy of the results is reliant on the quality of the video tape player being used and the counter information that had been recorded onto the video tape. This accuracy is usually within 0.1 seconds of the event that had occurred.

There are many inherent disadvantages with an analog system of this type. The most obvious is the limitations of the actual media. It is impossible to have random access to any location on the video tape instantaneously. Video tape is a linear medium that can take several minutes to locate a desired piece of video. This is one of the major limitations in using video tape for research purposes.

It is very difficult to make comparisons between different subjects. The researcher typically watches one subject then winds the tape onto the next subject while trying to remember the details of the first subject for making comparisons. Even with a second video tape, the two are not visible at any one time. A possible alternative would be to have two video players synchronised together. This is an expensive option that is not realistic on limited research budgets.

In many of the situations where the researcher is analysing multiple subjects, they may rely on their notes from the previous subject. This makes the results less accurate since the researchers notes become the source material that is often referred to, and compared against, instead of the original video clip.

Less time spent analysing the video tape information and taking notes, means more productive time is available for analysing the results. Rewinding tapes to different locations, together with the inherent time delays of the video tape player's head mechanisms all contribute to wasted time. Another area of lost time for the researchers is re-entering results taken, into a computer, for further analysis purposes. Entering the results into the computer as they are found would be an advantage for the later stages in comparing results. If the process is difficult then the researcher may not bother checking their results or doing a thorough job.

1.3 Other Applications of Time Based Analysis

Research into time based information is not confined to the analysis of human behaviour but crosses many disciplines of research. In the area of physics, high speed laboratory experiments need to be examined in detail to find out what reactions occurred. The human eye and brain cannot process high speed information so the experiments are usually recorded onto video tape and then played back in slow motion. In the field of Botany, researchers may wish to examine plant growth occurring over a long period. Time lapse photography by a camera every couple of hours yields the desired information. Playing back the sequence of images, shows the stages of the plant's growth in a short period.

The process of analysing time based information has commonly consisted of using a video tape player with pen and paper. The central reasons for using video tape have been the ability to capture visual information at a relatively inexpensive cost by using video tape players and media.

1.4 Conclusions

To better understand what researchers of video tape information are faced with, the available methods for analysing time based information are examined in chapter 2. From this, the benefits and disadvantages of the available options can be weighed and considered in the light of current technological advances in the digital video area.

This thesis is directed at exploring whether digital video can be used to assist in the analysis of time based information. The design of key components of a system are explored together with the interaction of researchers with a low cost digital video based analysis system. The use of digital video has not currently been used in the design of an analysis tool to assist in the exploration of human behaviour. The design and interaction of a prototype system on an Apple Macintosh computer, known as *Video Transcriptor*, with researchers will be examined to see how it can benefit in the analysis of time based information. Other features of digital video technology are also exploited to see how they can improve the quality of research carried out.

Chapter 2

Time Based Analysis Systems

> *"Video is widely perceived as an important medium for psychological research, because video recording makes the fleeting particulars of human interaction available as data for detailed analysis, while retaining much of the context of the event. Though the benefits of using video data are high, the process can be prohibitively time-consuming."*
>
> \- Jeremy Roschelle [39].

This chapter presents an overview of the different systems and techniques available for the analysis of time based information. The main emphasis is on computer-assisted systems that analyse video-based information, although traditional methods are discussed.

This chapter begins by examining observational methods used for analysing time based information. Following this is a survey of past and present video analysis techniques available and then a look at computer-assisted systems. Each system's strengths and weaknesses are discussed.

2.1 Observational Methods

> *"Most of the data in applied behaviour analysis have been and still are collected by observers using various paper-and-pencil methods."*
>
> \- Alan Repp [37].

Observational methods using pen and paper will always be favoured by many researchers. The main attraction is the simplicity of the procedure and the inexpensive overheads. Even though this method may seem simple to the researcher, the results and productivity achieved can be less than satisfactory. Transcripts and coded databases lead to more organised note-taking procedures but the data collected from these improved methods is still cumbersome for the extraction of the required information.

The simplest method available for a researcher is to take observational notes as each event occurs. Under certain conditions this approach may suffice. These conditions normally apply when the time based information occurs at reasonable intervals without the possibility of event conflicts occurring that may cause the researcher to lose information. Typical scenarios for examining time based information will have many independent events occurring randomly. In many situations these random events will occur simultaneously leaving the observer unable to cope with the multitude of events that need servicing.

The observer themselves may be visible or removed from the actual subject of the time based information. This would depend entirely on the nature of the information being examined and any outside factors that have to be taken into account. For example, a person counting the flow of traffic over a period of time would not care where they were located as long as they could see all the traffic flow. By comparison, an observer of young children, interacting with each other, may wish to be removed from the subjects. In this case the observer may be located behind one-way glass in a special experimentation laboratory used for this type of analysis.

The nature of the notes that the observer takes will be appropriate to the type of research being undertaken. For example the observer may be wanting to examine the frequency of events occurring within specified time periods, or alternatively the lengths of individual events as they occur. The notes taken will be dependent upon the type of analysis required. Each observer will have their own methods of note-taking and their own shorthand. This is independent of the information being analysed.

With observational analysis techniques it is difficult to confirm whether the results are reliable. Once the event occurs during the session there is no way for a single researcher to check if the results obtained were correct. This makes it very difficult to accurately make any conclusions if there is any discrepancy in the results. The researcher cannot re-examine the event again; all they have is their notes. To improve the reliability of the results, multiple researchers must be employed.

To avoid the problems associated with event collisions, multiple observers can be employed to specialise in observing one specific event type. This allows complex behaviour patterns to be successfully observed with the minimum of equipment. A single observer can only handle a limited number of events occurring within a single time interval.

All experiments have large costs no matter what type of scheme is used; people's time, the cost of rooms, planning, and the analysis of the results. It is not just the cost of pens and the paper, which is very small. Other techniques, examined later, might be more cost effective than using observation techniques.

Other techniques of note-taking include the use of transcripts and coded databases. Transcripts consist of a written log of everything that happens during the given time period being examined. The researcher then writes notes in the margins of this transcript. Thus annotations conveniently appear next to the transcript to which they refer. This method is more appropriate for use after the session where the transcript has already been created. It cannot be used directly when observing.

Coded databases on the other hand are directly relevant since they can be done while observing. A coded database is a shorthand notation for specific events that occur. For example, a database can be set up of specific events that are to be examined in detail. Whenever an event occurs the code for the specific event is taken down instead of having to write a longhand description about each event.

Depending on the nature of the source material, observational note-taking methods may not be the best approach available. Observational techniques may not be appropriate to use for chemical reactions that occur in a split second. The human eye and brain may not be able to pick up everything that occurred necessary to make accurate observations. Under this situation, a high speed video camera recording the chemical reaction may be better suited. The video could then be analysed frame-by-frame to obtain the results.

At the other extreme, very slow time based information can produce similar problems for an observer. Trying to analyse a plant's growth is as difficult as trying to examine the spontaneous reactions in the chemical experiment. The plant's growth is too slow for the eye and brain to easily perceive changes over long periods of time. Using time lapse photography or a video camera pre-set to record a static video frame every few hours would be a better situation. The resulting series of photographs or video frames can easily show the progression of the plant's growth over a period of time.

It can be very difficult to extract information from hand written notes or coded databases. The observer must go through his or her notes in detail whenever information is to be extracted. This is one major limitation in pen and paper based note-taking methods.

It is worth noting that observational analysis techniques of this kind can offer some advantages over video tape systems, discussed in the next section. It is generally faster to retrieve a location in a transcript than it is to find a location on video tape. The annotations associated directly with the transcript offer a better direct relationship to the original source material. Live observational techniques allow the researcher to take down additional notes about events that may occur that would not be captured onto video tape. For example, two children in the corner of a room fighting, out of a video camera's field of vision, might have a direct influence on the subject being viewed. The researcher would be able to pick this up immediately while someone who looks at the video tape results would be unable to since they would have no knowledge of the two children. The direct observational technique also forces the researcher to think while events are occurring. They are able to instantly ask the subject a question as to why they might have done something. Video tape technology does not offer this feedback when viewing it at a later date.

2.2 Video Tape Systems

A different group of time based analysis methods incorporate the use of a video tape player. Video tape is capable of preserving vocal intonation, gestures, and facial expressions of interaction and communication that are difficult to capture through conventional observational techniques or to represent in traditional data records.

Video is a medium that cannot be easily annotated. Marking and logging of video must take place on other media such as paper or computer records. There is no easy way to "write in the margins" of a videotape, so that comments on an edit decision or opinions on the subject matter can be easily passed on to others. In addition, copying material with video tapes in the same or different formats can result in significant quality loss. Working with video tape media is cumbersome and time-consuming in comparison with the use of more traditional media such as transcripts and coded databases. Since the introduction of video tape technology many different schemes have evolved to assist in the analysis of time based information. Video tape players and media are inexpensive and readily available making it widely accessible for observers as a tool to assist them.

Using video tape allows the observer to change the emphasis of having to note every behaviour or event as they happen to a situation where they can review the source material repeatedly without losing any of the original information. This is an advantage to the observer who can now perform more complex observations on the information available.

Another major advantage is the ability to record time based information that was either too fast or too slow to observe successfully with the naked eye and brain. Now the observer examine what they previously were unable to see.

Video can also be used to assist in examining children's behavioural characteristics, as a group of pre-schoolers interact with each other. A vast amount of events occur simultaneously in this environment. The information can be processed by observational techniques if enough observers are available to deal with all the required events being analysed. Realistically, alternative methods may be better suited in this scenario. By recording the interaction of the children onto film or video tape, a single observer is able to examine the contents of the tape in more detail and in a relaxed atmosphere.

The observer may play the tape back for each of the many behaviours to be examined. If there are any controversial sequences in the video then it can be played back in slow motion or even paused at any particular place. There is no problem in checking for errors in the results since the original information is available to be checked again, and even by another researcher if necessary.

Several things that video tape are unable to record are background information to the visual information. This includes events leading up to the current session and also things that are happening out of visual sight of the video camera. Events that lead up to the current situation are a problem for any analysis method but can affect the content of any subject being analysed. Additional notes must be taken by the researcher if this type of background information is of significant importance such as the names or backgrounds of subjects and events that lead up to the session being analysed. Events occurring out of visible sight are also a problem but observational methods may also suffer this problem. These problems face all analysis methods and cannot be easily overcome.

Several inherent problems with video tape technology exist that are of concern. The main problem associated with the in-depth analysis of video tape is the time required to search for information. Since video tape media is linear, it takes several minutes to physically locate a section of tape that is required. This is a significant problem that hinders the use of video tape in many instances for making cross comparisons. Other problems exist, such as video tape media deteriorating over a period of time causing the information to be rendered less useful. Constant use of the video tape can degrade and stretch the physical tape. Researchers analysing video tape information normally employ techniques that often cause excessive stress and wear on the physical tape. These techniques include the constant starting and stopping of the tape, pausing the tape for extended periods of time, fast forward and rewinding the tape, and stepping through the tape. Under normal circumstances tapes are normally played back at normal speed, but when trying to analyse the contents of the video tape there is stress placed on the tape, shortening the life span and the playback quality of the tape.

Jeremy Roschelle [39] found that observers using video tape as their primary source of information encountered many problems. The observers' main difficulties were with annotation and retrieval of information. This forced many researchers to work with secondary data such as transcripts or codes, rather than analysing the video itself.

Trying to associate the annotations with the source material is one of the major problems encountered when using video tape as an analysis media. Under observational analysis techniques the observer takes annotations on the same paper as the transcript of the events that occurred, and can easily find its location in coded databases. With video tape media this is not possible without any further assistance.

2.3 Computer-Assisted Systems

Computers can be used to assist in the analysis of time based information. There are three main ways for the computer to be used; as a stand-alone note-taking tool, a combination of note-taking and controlling of the video tape player, and as a digital video analysis tool.

Many types of computer-assisted systems have been developed to assist in the analysis of time based information. Many of these systems offer advantages over standard video tape based systems, but do not provide any final solutions. An overview of the different types of systems follows, categorised into three main types; computer based note-taking only [16, 37], computer and audio or video tape [8, 11, 13, 25, 31, 39, 43], and entirely computer based [33].

2.3.1 Computer-Analysis From Hand-Collected Data

Alan Repp et al. [37] introduced a data collection and analysis system based on an Epson HX-20 portable computer. This system offered the ability to record multiple behaviours in real-time.

The system was based on a computer program using an Epson HX-20 portable computer that provided continuous and simultaneous recording of up to 43 behaviours. The researcher can hold down any number of the 43 buttons, previously assigned for a particular behaviour, to record the start and end times of each behavioural occurrence.

The system offers benefits over the restrictions found in pen and paper methods. These restrictions include event recording, duration recording and other various time-sampling procedures. Event recording is inappropriate for behaviours whose duration varies. Duration recording on the other hand becomes unwieldy when several behaviours are recorded. The other various time-sampling procedures inaccurately reflect duration and/or frequency of behaviours sampled.

Other systems of a similar nature had several limitations when compared to this system. This system was cheap, portable and could encode more than one behaviour at a time. The software is self-contained in features, recording frequency, duration and the sequence of each of the behaviours. It also provides analysis functions on the data recorded by the observer in the form of complex reports. The reporting features show the order of each of the responses that occurred, the session total of all behaviours, the number, rate, duration and percentage of each response occurring. It also provides an automatic means of assessing inter-observer agreement by comparing different observers' results. This is necessary to find out how accurate the results that have been calculated are.

The major advantage of this system is that it saves time in the recording of behaviours once the observer has learnt how to use the system. It provides the ability to code multiple behaviours at the same time which reduces the time needed to record the behaviours. Since all the information is in an electronic form, complex analysis of the data can be performed to find out different types of behavioural characteristics easily. The major disadvantage is the user being required to remember what each of the 43 buttons are recording. When simultaneous events occur, the user has to remember and locate each button necessary, resulting in delays to the times that the events are recorded.

2.3.2 Lap Computer

Moshe Guthertz et al. [16], in 1989 introduced a portable hand-held data collection device to assist in behavioural analysis of mother and infant relationships.

Laboratory and field observation techniques that normally employ pen and paper techniques use time-sample unit methods to code occurrences of behaviours. This was limited to recording behaviour frequencies and not durations or behavioural sequences.

Systems had already been developed using keyboards and tape recorders for recording duration and sequential information like the Digitorg system [13], discussed in Section 2.3.4. This research began using a small portable recording device, the Tandy-102 from Radio Shack, that could download its results onto an IBM-XT computer for later analysis.

The experiments carried out involved the continuous coding methods of mother and infant interaction behaviours from video tapes taken in the laboratory. The lap computer coded the video tape behaviours by the observer pressing numerical codes in the keyboard which represented each behaviour. The video tape itself was time marked second-by-second using a time lapse generator during the video taping. The video is rewound back to the time zero location and both the lap computer and the video-tape are started at the same time. This method does not involve any interaction between the lap computer or the video tape thus the results are only accurate to the nearest second or more. The coding process involves coding each subject one at a time. Once the mother's behaviour has been analysed the video tape is rewound and the infant's behaviour is analysed.

Once all of the behaviours have been coded, the results from the lap computer are downloaded using the RS-232 serial ports and communications software for the transfer. The results are then analysed using software on the IBM-XT computer.

2.3.3 GALATEA

GALATEA, developed at the University of Chicago by Futrelle et al. [11] in 1975, was an early computer-aided analysis system designed for rapid extraction of data from 35mm film using interactive graphics techniques. It was a pioneering system incorporating the use of computers with film data. The user would indicate important features on the film by using a digitising pen to follow an image on the screen. The user received feedback from the system in the form of an animated computer generated movie overlaid on the original film. The relationship between the relevant components can be seen in figure 2.1.

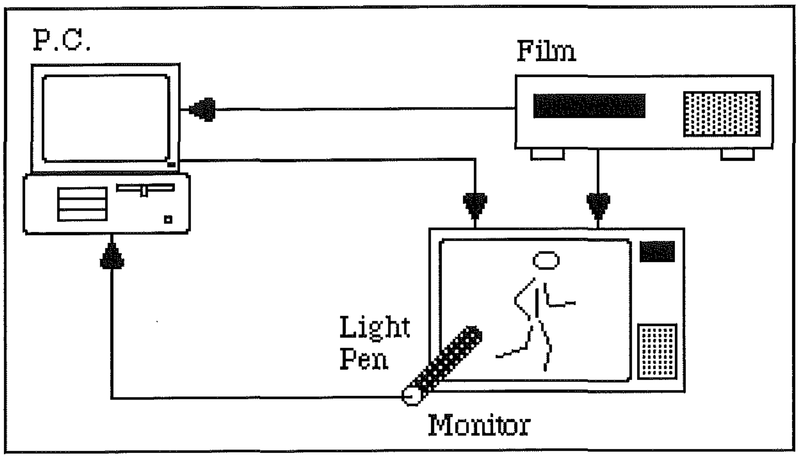


Figure 2.1. GALATEA's Data Entry Technique.

This method allowed the user to record complex, naturally occurring event sequences or processes, storing the data for later analysis. The information entered by the digitising pen was the only data that the system saw and was recorded in its internal data structures. GALATEA could generate animations of movements or changes in shapes that could be viewed from any angle, and could produce a number of calculations based on this data. GALATEA was an extraordinarily powerful tool for its time, the mid seventies, anticipating many advances in image processing and interactive graphics.

Not only was GALATEA the first in its field but the designers were well aware of more complex issues that could be employed to further enhance the process of analysing time based information. For example, they were aware of digitising individual video frames and even using the computer to perform numerical pattern recognition to identify items of interest. This process could even extract the numerical parameters similar to those collected by the current digitising pen in GALATEA. The amount of information required for this was a major concern, which limited further research. Research into this was never carried out in depth due to time constraints in order to develop the already new and difficult GALATEA system. The working system GALATEA was developed and run on a PDP-11 which had limited memory and disk storage. Implementing anything coming close to what was discussed by Futrelle et al. was unrealistic for the period but shows what type of system they were thinking of just under twenty years ago. Over the years there has not been a dramatic change in systems of this nature. Instead of controlling film the newer systems control video tape machines. The analysis software has been the only major development during this period, until the recent wide-spread availability of digital video standards and resources on personal computers.

2.3.4 Digitorg

Digitorg (Digital Behaviour Organ), developed by Clifton Lee Gass [13] in 1977, was an inexpensive, portable, field-operable event recorder. It was used to record behavioural and spatial information of hummingbirds behaviour in real-time onto audio tape.

The system comprised of an encoding keyboard and a tape recorder for the observer to use. Behavioural information was recorded by pressing buttons on a keyboard for each type of behaviour to be analysed. Spatial information was recorded by touching a magnetic wand to a paper map overlying a matrix of switches. The signals from the switches were digitised for recording in binary code on a single track of the stereo audio tape recorder. Additional voice notes could be recorded onto the second audio track on the tape.

The encoding keyboard consisted of 256 software defined codes. There is one major restriction with the system, only a single event could be coded at single time. This was because the system is insensitive for about 70 msec after each switch closure due to an electrical limitation of the system. The 256 switches available could be configured for recording spatial information and as behavioural coding keys.

The encoding keyboard and the tape recorder comprised the data capture component of the system. The analysis of the data recorded was carried out on a separate PDP-11 minicomputer. The tape containing the behavioural and spatial information had to be played back at four times its normal speed through an interface into the PDP-11 which read the information in as a stream of digital information. This process recorded the time and identity of each event that happened.

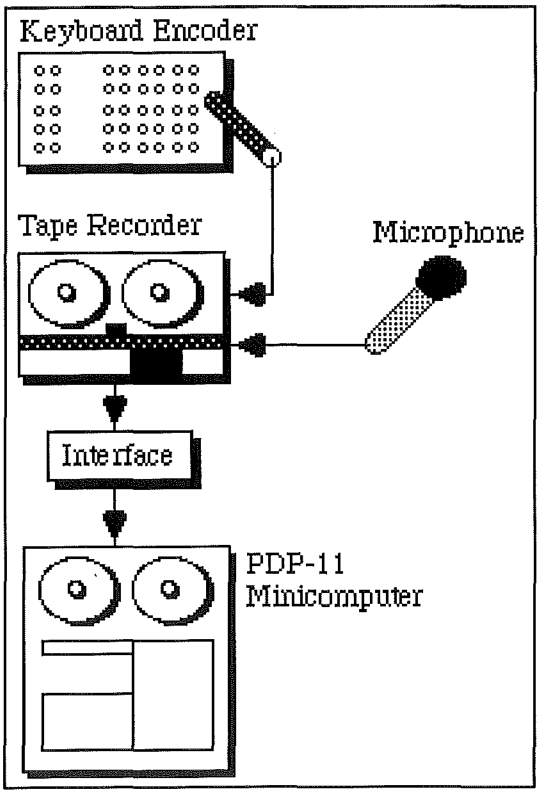


Figure 2.2. Digital Behaviour Organ (Digitorg).

Figure 2.2 shows the two inputs to the tape recorder, one from the encoding keyboard and the other from the microphone. The coding stage of the system could be done anywhere that the tape recorder could be moved to. Digitorg could be considered a portable system for its time. When the encoded results were to be analysed they had to be played back through the interface into the PDP-11 minicomputer.

The main limitation of the Digitorg system was the accuracy of the recorded information. Digitorg was mainly used to record behaviours of subjects in a live situation. The observer had to encode specific behavioural events by using the keyboard while also recording spatial information. Depending on the type and amount of information, this could become a burdensome job. The recorded events would not be precisely accurate. Once the event occurs, the observer has to first of all work out what it was, what button or what location to use the wand with, and then actually encode the event. This whole process is not instantaneous so the results would have varying degrees of accuracy.

2.3.5 VideoLogger

VideoLogger, developed at Columbia University, New York, by Robert Krauss [25] in 1988, is a computer-assisted system permitting multiple events to be coded from video tape. It was developed using a video player, an Apple II computer, a John Bell 6522 timing card, and a pair of switches.

VideoLogger consists of five software programs; one to make the timing signals on the video tape, one to record the onset time of an event and the duration of the button press, one to read and print the coding data from the disk, one to operate the timer and finally the last program to process the compiled code used in the software package.

The audio tracks of the video serve two purposes; the first contains the normal audio associated with the video and the second containing timing information. By using one of the five programs of VideoLogger, the user plays the video tape back and marks the start and end of the segment to be coded. The video tape is then rewound and played back. When the computer detects the starting mark on the second audio track the program for recording events is activated, the timing circuits of the 6522 card are activated and an audible cue is sounded for the observer to begin recording events. Only one event type is recorded on each pass of the video tape. The observer presses one of the buttons to record the start and end times of any number of events. All button presses and releases are recorded into memory until the end marker is reached on the second audio track of the video tape. The data that was entered by the observer is then saved to disk. This task is repeated for each event type that is to be recorded saving each of the results onto disk when finished. Once all the different events have been recorded the program for reading in the results is run and this can be used to analyse the results and print out the desired results.

VideoLogger has the capability to record simultaneous events that many previous systems had trouble handling. This is achieved by recording each of the simultaneous events on a different pass over the video tape. It has an internal limitation of only allowing 256 events of the same type to occur during the same period. This can be overcome by just making two shorter analysis periods on the video tape.

2.3.6 Experimental Video Annotator (EVA)

EVA, designed by Wendy Mackay [6, 31] in 1989, was a system that was born out of frustration with the prospect of having to analyse hours of video tape information (figure 2.3).

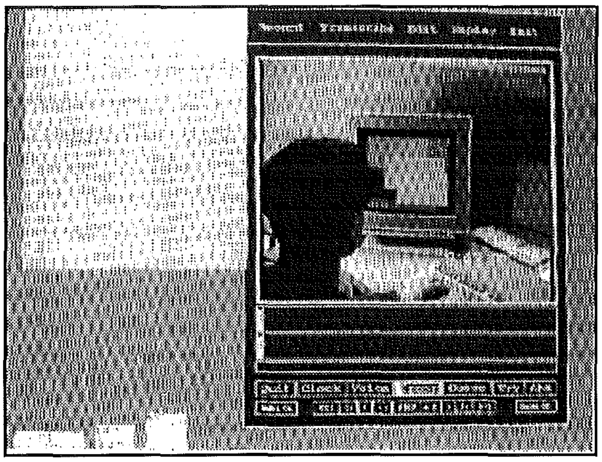


Figure 2.3. Experimental Video Annotator. (From [6]).

EVA runs on a multimedia workstation developed by the Visual Computing Group at MIT's Project Athena. The hardware consists of DEC MicroVAX or IBM PC/RT workstations, running UNIX and the X Window System. Each workstation has a Parallax board that is capable of digitising video in real-time and adds support for colour graphics. EVA is written in Athena Muse, a language for creating interactive multimedia applications. Athena Muse was an eight year, US$100 million experiment in computer aided education [32].

Wendy Mackay [31] states "Researchers need to capture and record their impressions of events as they occur. Taking notes is helpful, as is marking an event for further analysis. Even though it is almost always necessary to review the video again, real-time annotation of the data can save a tremendous amount of time."

EVA is a prototype system that runs on specialist hardware and controls a video tape recorder. Prior to using the system, the user can devise a set of buttons indicating different events to be examined. As the events occur the researcher can press the appropriate button to record the location for later analysis. Another capability is to annotate the data symbolically. This consists of digitising static images, as icons, so they can be used as references for the researcher. A small colour icon is more meaningful than the time that the event occurs for the researcher for later analysis.

EVA was primarily used to evaluate new software packages by subjects. During the session the subject sits in front of a computer and begins to use the new software package. A video camera is directed at the subject's face or the screen of the computer that the subject is working on. The researcher sits in front of the EVA system in the same room or in a different room. The live video appears on the screen of the researchers computer and is simultaneously recorded to video tape. As events occur the researcher can use the buttons to tag ones of interest and to create small icons of important events.

The annotation during the session saves time later when the researcher is ready to review and analyse the session. The researcher only needs to go back to the tagged sections of the recording for further analysis. Using the small icons that may have been digitised makes it easy to see what the event is without having to rewind the video tape. During the analysis phase the researcher controls the video information with video tape player style controls. Annotations are performed with a text editor that links the record to the tags created. At any location in the video, a multimedia edit log is created where notes, snapshots of video, and other important information can be recorded. This process builds up a random-access database of information for the researcher to analyse.

EVA is an important system that provides a new approach to analysing video information. There are many advantages with this system making it easier and more productive to use than the other systems analysed. However EVA requires specialist hardware to run that may only be available in the most well equipped labs [6]. It is also limited once again by properties of video tape. The access of information is slow for reviewing but is compensated somewhat by the multimedia edit log. The act of tagging and annotating events can prevent the researcher from actually concentrating on the events themselves, thus events can be missed or even tagged late.

2.3.7 C-QUAL

C-QUAL, developed by Duncan et al. [8] in 1990, is a system for the computer-aided transcription of action sequences from video tape. It is a direct descendent of GALATEA, an earlier system developed at the University of Chicago by Futrelle et al. [11], from which concepts and techniques were adopted. C-QUAL is a system designed for research on face-to-face interaction involving speech and body motion.

Generating accurate transcriptions is problematic for three reasons. The first is that manual transcription is an arduous and time-consuming process. The observer has to record not only the start and end times, but also its sequential location with respect to all other events. The second problem is the integration of speech and body motion. Aural information must be combined with the visual data for it to be of use. One cannot rely on the video tape to obtain video information since playing back the video tape at different speeds does not result in the audio being played back in a useful condition. The third main problem is the entry of the data into the computer database. This process is normally slow and tedious and subject to clerical error.

The original video of the subject is copied onto a second generation tape with machine readable frame numbers and speech representation inserted. The machine readable frame numbers are inserted using a time-code writing device while the audio representation is created using a Sequel-Analysis Module (SAM-259) device.

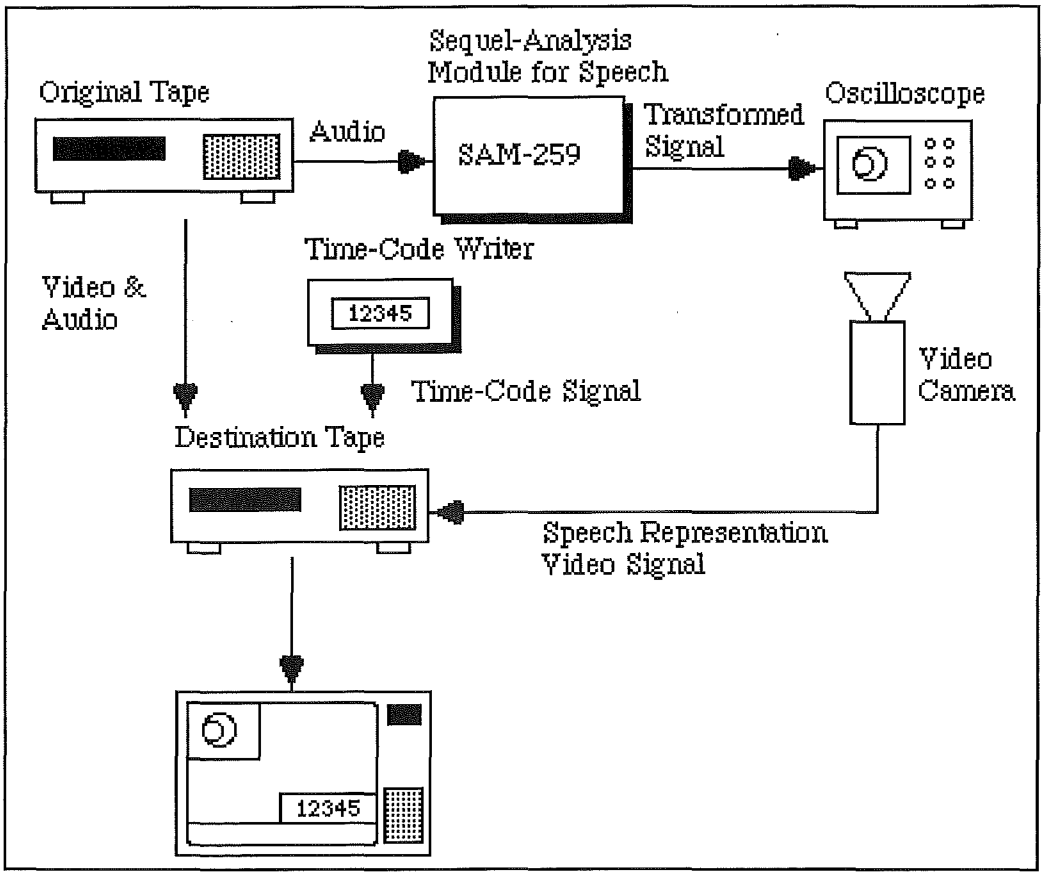


Figure 2.4. Set-up of The Source Video Tape For C-QUAL.

Figure 2.4 shows the process of creating the second generation video tape with the additional information. The audio information from the original video tape is fed through the Sequel-Analysis Module and the modified signal is displayed onto an oscilloscope. The display of the oscilloscope is videoed and mixed with the original video signal onto the second video tape positioned in the upper left hand corner. The audio information is recorded directly onto the second video tape. An additional time-code signal is written onto the tape so that the computer can read the video's precise location when the observer is analysing the video.

With the audio information being recorded onto the new video tape in a visual format, it is possible to examine the audio information at any speed. The audio information in C-QUAL can be represented in three ways; the original audio track, the SAM based representation, and a text-based representation, making it relatively easy to analyse audio information.

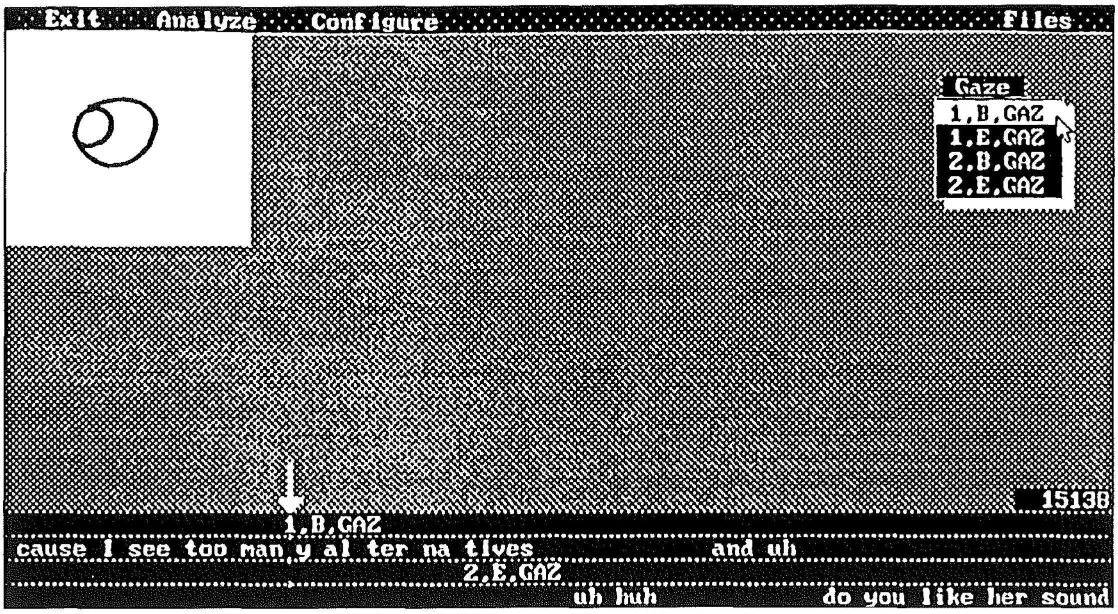


Figure 2.5. C-QUAL Transcribing Screen (Video image is omitted. Sequel-Analysis Module display is simulated), [8].

Figure 2.5 shows what the observer sees as he or she is analysing the contents of the video tape. The original video source is displayed on the screen with additional information that had been mixed with it. Note that figure 2.5 does not show any video image. The grey area is where the video signal would be shown. In the top left hand corner is the output from the SAM unit. This illustrates certain properties of speech. The researcher is able to see the shape of sounds made indicating the stress, pitch, and phonetic structure of sounds made by the subject. Figure 2.6 shows examples of what the SAM module produces as output for certain sounds.

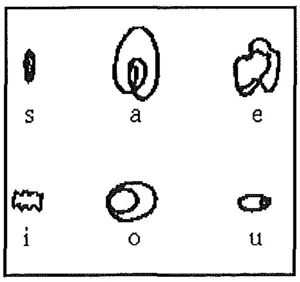


Figure 2.6. Example Speech Representation Produced by the Sequel-Analysis Module (SAM-259).

The lower right hand corner of figure 2.5 contains the time-code information from the video. The lower portion of the screen is the text-based speech representation. This is extracted directly from the computer and displayed. It can be changed by the observer as he or she is analysing the video tape.

As the researcher examines the tape, he or she enters the beginning and end points of an event. The time-code from the video tape is read and stored with the researcher's transcribed information onto disk. The transcribed information can then be displayed back simultaneously while examining the video information in the future. The time-codes that have been recorded with the information on disk are used to link the video tape information with the notes taken. When playing back the video tape the transcribed information from the computer is fed directly to the video monitor together with the video information.

C-QUAL is a very comprehensive transcription system with a powerful set of features. It offers the ability to locate any section of the video with little effort. The computer-controlled video will automatically scan for the correct location based on the time-code information. The transcribed notes are automatically linked with the video as it is displayed which makes the analysis stage very simple. The transcription notes are restricted to two categories. The results of the transcribed notes are exported into a separate application called THEME for further analysis to take place.

The major problems with this system is its complexity and high cost in setting up all the necessary components. This system is not designed for anyone but large research institutions with a large budget to spend. The main expense for the system is on the video timebase corrector, valued at approximately NZ$17,000. The features make it an attractive video-based system but the cost makes it inaccessible to all but a few.

2.3.8 VideoNoter

VideoNoter, developed by Roschelle et al. [39] in 1991, represents a well integrated environment for analysing time based information. VideoNoter supports annotation and retrieval of video data and offers the presentation of multiple analysis of the same data.

Video tape is the primary source of information with the secondary records such as transcripts, codes and annotations linked by the timing information held on the video tape. Multiple analysis of the same data is supported and can be viewed side-by-side to compare how different observers responded. All transcripts and analytic codes are linked to the video thus the retrieval of information from video is relatively easy, without the need to preview the tape. VideoNoter controls the VCR so it can move the tape to the desired location. Without it, one must guess how long to fast forward or rewind the tape.

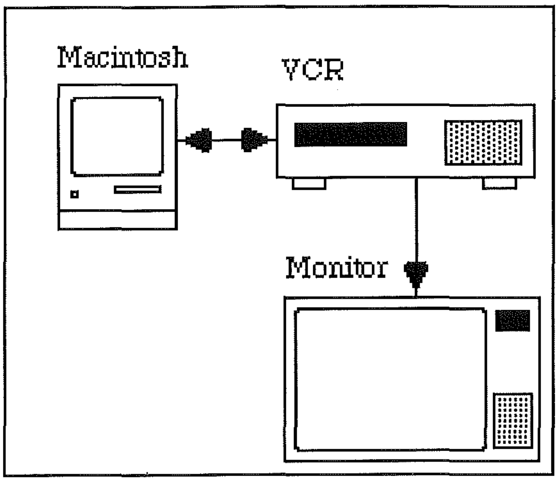


Figure 2.7. An Overview of VideoNoter.

Figure 2.7 illustrates how the computer controls the video player. The notes are displayed on the Macintosh's screen while the video and audio from the video tape is displayed on the monitor attached to the video player.

VideoNoter operates with the user using a "progressive refinement" technique. By starting with coarse coding of the entire video and then gradually adding layers of more detailed annotations, comparisons are easily achieved. Since all the notes are stored on the computer, the locations on different tapes are automatically known. The two locations of events can be found without trouble except for the limitation of searching through the tape to find the requested locations. The main limitation with VideoNoter once again comes from the video tape player.

2.3.9 CAMERA

CAMERA, developed by Maarten van der Vlugt et al [43] in 1992, is a hardware and software system for coding complex behavioural interactions from video tape. It is a system that was designed to improve accuracy, reliability, and training standards in the coding of behaviour. The main reason that CAMERA was developed, was that the coding of behaviours was not reliable and contained errors. The sources of these errors arise from two areas; the observer and the coding instrument.

CAMERA was designed to improve the reliability and accuracy in the coding of behaviour.

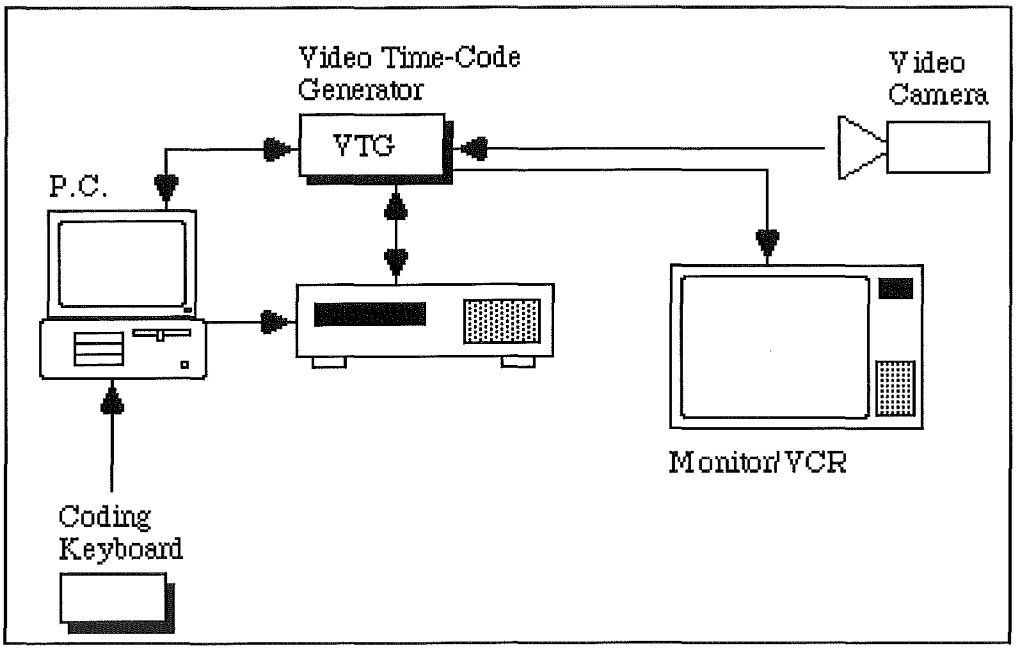


Figure 2.8. Main Components of the CAMERA System.

Figure 2.8 illustrates the interaction of the different components of CAMERA.

The system writes a time-code onto the video tape for each frame. This code is automatically retrieved and checked when scanning through the tape. When an event occurs, the appropriate button on the event keyboard is pressed for the duration of the event. This codes the start and end times of the event, together with the mnemonic label from the event keyboard, into the memory of the PC. Each button on the event keyboard has a specific meaning that maps to a type of event that is being checked for. Whenever a button is pressed on the keyboard the user is presented with visual feedback on the monitor together with a unique pitch for auditory feedback. The mnemonic label is displayed on the video monitor while the event occurs. If multiple events occur then the labels are displayed side by side on the video monitor. To code all the different events, it is common to concentrate on a single event type each time the source is analysed. The coded files of the key presses are then merged into one.

Improvements in the quality of the coding process are achieved by having several independent people analysing the same information and then comparing their results. Maarten van der Vlugt [43] observed "This feature appears to be valuable not only for judging the accuracy of the coding, but also for training new observers in a short time and without intensive guidance."

This performs two functions at once. It detects events that may have been missed by one of the observers and can highlight certain types of events that may not be considered important by the trainee observer but the trainer finds important. By comparing the two results one can see how well the trainee did and can be used as a basis to check the accuracy of results in other experiments that may have been done in the past.

2.3.10 Monitoring Animal's Movements Using Digitised Video Images

This system, developed by Richard Olivo et al. [33] in 1988, at Smith College, Northampton, Massachusetts was one of the pioneering techniques of using digital video based techniques to analyse animal's movements.

The system could be run on either a Macintosh II or a PC/AT. It used a frame-grabber to record a black and white image of the animal. Each image recorded was 640x480 pixels for the PC/AT and 512x480 pixels for the Macintosh II computer. The frame-grabber cards were reasonably expensive, costing US$1,000 for the Macintosh II and US$1,500 for the PC/AT computer.

The idea of the system was to take multiple frames of the animals to see the movement of the image by taking the differences in the pixels in the images. The tests performed were using creatures such as crayfish that were held down on a table. The principal function of the experiments was to see the motion of the crayfish's legs and head. The time taken to digitise each image was substantial and the image quality was not very high. This was only intended for very small scale experiments with items that exhibited small movements.

This was a pioneering system using new technologies available at the time. It is not appropriate for analysing large amounts of information like that required by previous systems that analyse human behaviour. It is however an important system that used innovative technology and ideas in its time.

2.4 Comparison of Analysis Systems

There will always be different needs and different situations for analysing time based information. In one instance an observer may only be interested in the number of people entering a shop while another observer may wish to examine every motion and every word spoken of each of those customer that entered the shop. An observer with a pen and a piece of paper would easily be able to cope with the first scenario. Trying to record every event that occurs with pen and paper techniques would become overwhelming unless there were dozens of observers all crammed into the one shop each analysing a different behaviour. An alternative system using video tape with computer control may be better suited to this particular scenario. In contrast the video tape system would be overkill in the first scenario.

No single system prevails as the dominant winner over the others. Each system offers some useful advances in features to the researcher but usually still have some limitations.

Pen and paper observational methods are good examples of easy to use analysis systems that can cope with relatively simple scenarios. They do not involve much expense and are simple to use and can achieve fast results with this method. One major limitation is large or complex types of information that have to be analysed. The type of information normally analysed is live whereby the events cannot be repeated. More researchers can help by splitting the original workload up into smaller units to overcome this. The second major limitation of pen and paper techniques is the correlation and analysis of the notes that have been taken. This is normally not a simple task. Quite often the data is entered into a computer where more complex analysis can take place. Under normal conditions the correlation of the data collected can be far more time-consuming than the time to take notes and to enter them into a computer. This in more apparent with more data that has been collected. There is also no way to check the accuracy of the results that have been recorded unless there are several people involved, each checking the others results.

Alan Repp's [37] hand-held system was basically a faster way of processing observational notes by supplying a complex event keyboard to enter event details simultaneously. Moshe Guthertz's [16] hand held system offered the capability of entering the data directly into a machine readable form. These two hand held systems offered more speed and improved reliability for the researcher over conventional pen and paper techniques. However they still do not offer the accuracy that is needed and have no way of checking the accuracy of the results taken short of several people coding the same information.

Video player systems offer more reliability than using straight pen and paper observational methods on complex time based information. By having the source material recorded, the observer is able to go through the information slowly and repeatedly until all the information has been analysed. There is no need to employ multiple observers since the one observer can examine the tape several times for each behaviour or event type being analysed. This assists in making more reliable observations. The limitations of the video tape methods include the time spent analysing the video tape. The control and speed of video tape players are slow and burdensome. The tape media is linear, so to find a particular location of the tape requires time for it to be located. The notes taken while analysing video tape media have the same properties as those taken with pen and paper techniques. They still have to be correlated and analysed further, possibly with the assistance of a computer.

Computer assisted systems offer more features for complex analysis requirements. The computer has been used in many ways to assist in the analysis of time based information. It has been used as a simple analysis tool of information already collected as in the Digitorg [13] system of the late 1970's. GALATEA [11] also from the 1970's used the computer to digitise the movement of objects which could later be analysed together with controlling film devices. The late 1980's and early 1990's saw many systems using the computer more to control video tape players and as integrated systems for note-taking.

There were two reasons for using computers with video tape players. Computers were becoming more affordable for the domestic market instead of large mainframe systems that large businesses could previously only afford. Video tape technology also just came onto the market in the late 1970's to early 1980's aimed at the domestic consumer. The combination of these two technologies has been increasing steadily with a significant number of systems available. The reason behind this is the price of video equipment continually dropping, while image quality, portability, and ease of use have all been improving. The same applies to computers, the price is continually dropping while the performance is continuing to double every year or two.

The use of computers controlling video tape players reduced the time spent for many researchers who had previously re-entered their hand-written notes into a computer for later analysis. The major limitations of these systems to many researchers are the cost of the special video tape players and the time-code generators necessary together with the normal limitations of linear video tape media.

Richard Olivo [33] offered a new approach to observing time based information by digitising sequences of images onto a computer to detect changes in movement over time. Although the image quality is not very high and only black and white, it was highly successful in the type of research it was used for. Although it was an expensive system and had its limitations in quality and usage, it could be seen as a predecessor of what has potentially become a very powerful analysis tool. It was only a couple of years later when digital video became standardised on many different platforms offering full colour, full motion quality at relatively inexpensive costs.

Digital Video has many new capabilities to offer time based analysis methods. With the technology rapidly expanding, standards being finalised, and the subsequent drop in prices of the necessary hardware, digital video has become a reality that needs to be explored in existing fields like human behavioural analysis.

Researchers who use video tape or digital video have many issues to consider in the area of data capture. Should the video camera be visible or hidden, fixed or moving, and how many cameras should be used. Should the researcher improve the aesthetic quality of the video by adjusting the lighting or should it be "as is"? Other issues in the analysis of the information is what portion of the video is considered important? Is it two or three key events in an hour and ignoring the rest, or is it five minutes of the tape in depth? What information is to be recorded? Will the sections of the entire video that are analysed concisely summarise the session or while it misrepresent it? These are all important issues that need some addressing with any method involved in capturing video and audio information.

Chapter 3

The Digital Video Revolution

> *"Still image graphics is to the 80's what media integration and dynamic data is to the 90's."*
>
> \- Joseph Ansanelli [2]

Video has been a major part of the public consciousness for six decades — since 1934 when Philadelphians witnessed the first full-scale television demonstration at the Franklin Institute [29]. The medium is an important part of our culture, and its influence on our lives continues to increase almost daily. It has seen significant changes in how people communicate. Since this time, television has become a universal communication medium in developed countries. Computers are becoming the next communication medium. Integrating the two technologies could have significant impact on the means of communication in society, as it will dramatically change the way in which we work and play. The main difference between the two technologies is that they are not directly compatible with each other, video data is analog while the computer is digital.

Both the invention of television in 1927 and the subsequent development of NTSC and PAL a number of years later, occurred prior to the development of digital technology. Although the vision of the designers of NTSC, PAL, SECAM and their variants was in many ways brilliant, the drawbacks of the analog nature of video are finally beginning to catch up with the medium. Technological limitations are becoming evident, particularly as the video and sound quality possible with digital technology improves dramatically.

There is a wide range of analog video standards such as NTSC, PAL, SECAM, PAL-M, PAL-N, and the proposed HDTV standards, including HD-MAC, and HiVision. There are additionally, proposed enhancements for NTSC, among them are SuperNTSC and also a number of HDTV formats such as Digicipher [29]. This variety of formats has an adverse effect of making program exchange more difficult. There is no accessible means for a typical video user to view any video format since there is no one global format in existence.

Each frame of analog video is noisy and blurred, not really suited for prolonged or critical viewing. Television was never intended for frame by frame viewing. It was a solution of providing moving pictures to the home in a time when electronic designs were quite different from what they are today [30].

In the past there have been insufficient standards in the computer industry for the storage, display, editing and manipulation of digital video data. The absence of such standards has resulted in few computer based applications with video capabilities. Those that do have video capabilities usually have minimal capabilities available. Furthermore, the absence of these standards in the past has also lead many software developers to customise their applications to specific hardware graphic and video cards. This means that the cost of producing such software is expensive and time consuming. With the introduction of standards for digital video, a wide range of products has become available for the consumer market.

In order for the two technologies to be compatible with each other, their signals have to be converted to the others' format. To convert analog data into a format that a computer can read, it must be passed through an analog to digital converter. Likewise for the computer's digital based information to be used on analog equipment it must pass through the reverse process of a digital to analog converter.

There are a number of digital video standards available, some for full-bandwidth video (such as D-1 and D-2) and some for compressed representations (CD-I, DVI, MPEG, Video for Windows, QuickTime movie format and many more). Some of these are proprietary formats.

Below in table 3.1 is an overview of some of the common analog and digital video formats available.

NTSC	current US and Japan analog video format
PAL	current European, Australian and New Zealand analog video format
SECAM	current French and Eastern Bloc analog video format
JPEG	still-image compression standard, can be used for video
MPEG	video compression standard for random-access media
MPEG-II	emerging standard for higher-quality compressed video
CD-I	consumer electronics interactive format
P*64	telecommunications video standard
D-1	current digital video production standard (component)
D-2	current digital video production standard (composite)
HiVision	Japanese proposed analog HDTV standard
HD-MAC	European proposed analog HDTV standard
Digicipher	General Instruments proposed US digital HDTV
SuperNTSC	Faroudja Laboratories proposed enhancement to NTSC

Table 3.1. Analog and Digital Video Formats. (From [29]).

A fully digital representation of video can bring many benefits, such as fast random access and high quality variable rate/direction playback. In addition, the digital domain allows easy image processing, manipulation, and image compression that the analog domain cannot offer. Digital representation is advantageous since it permits the integration of video as a data type along with all other digital data stored on computers. It also offers no loss of quality even after duplication or repeated use unlike analog methods of storage.

Before discussing digital video in detail the differences between analog and digital information will be examined. Analog technology is very susceptible to noise affecting the original signal. Since analog information is based on a direct representation of the original signal it is easier for corruption to occur.

We can examine the wave form of a sound to see the effects that noise can have on it. First of all we will represent our complex audio or video information as a simple wave form. This wave form repeats itself over a period of time and has a static amplitude. In real situations both of these variables will be changing frequently. As the frequency of a sound increases, the length of a single wave gets smaller. Similarly when the sound gets louder the wave form's amplitude gets larger. The point of this is to show that there is a direct relationship between the changes to the analog signal and the original sound (figure 3.1).

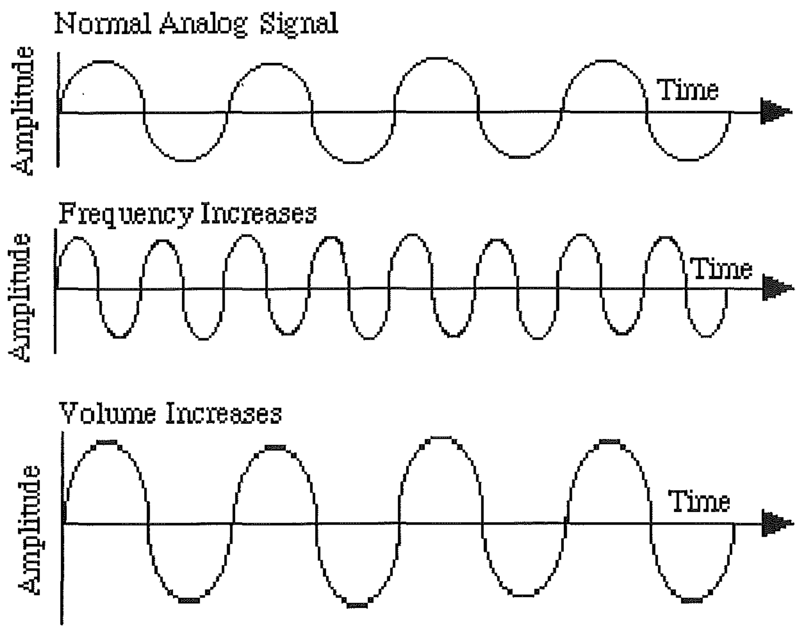


Figure 3.1. Properties of Analog Wave forms.

In contrast digital information normally has very little relationship to the original signal. The original signal gets transformed into a series of binary on/off signals. By increasing the frequency of the volume or the sound there is no direct relationship with the representation of the digital signal.

It is very easy for an analog signal to become corrupted and subsequently impossible to restore the original information. Noise to the analog signal can be introduced from many sources. In the case of video tape media one source of noise may be from natural ageing or from excessive wear in playback.

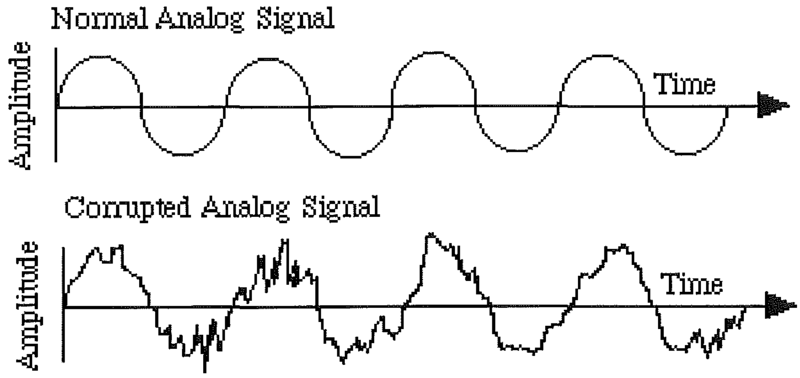


Figure 3.2. Corruption of an Analog Signal.

Figure 3.2 shows an example of noise introduced to an analog signal. The top wave form represents the original signal. The bottom wave form represents the signal after being subjected to severe noise.

It is impossible to reconstruct the original analog signal after corruption. Any variation in the original signal can have a dramatic effect on the original meaning. In audio information, the output can sound distorted and noisy. Likewise in video information, snow is introduced to the picture.

In digital information the original digital signal is easier to reconstruct, after noise is introduced (figure 3.3).

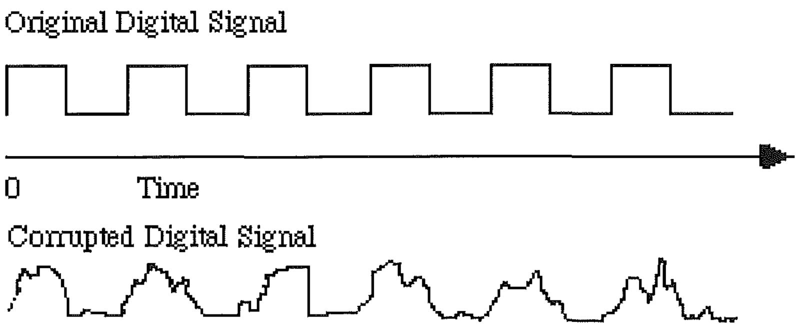


Figure 3.3. Corruption of a Digital Signal.

It is relatively straight forward to regenerate the original signal from the corrupted signal without significant error. A digital signal has only one of two possible values, a high or a low signal value. From figure 3.3 we can quite easily see what appears to be the high and low signals even after drastic corruption of the signal.

Other properties that analog information suffers from include the loss of fidelity in the video and audio each time it is copied. Each new generation is significantly different from its predecessor. It loses more of the original information each time it is reproduced, in comparison to digital information, where no information is lost. Natural wear of video tape also has dramatic effects on the quality of tape media. The electromagnetic patterns on analog tapes are a continuous stream of pulses. Over time the recorded patterns lose their strength and drop away. Digital recordings, in comparison, are streams of on/off pulses, so they can weaken over time and still faithfully reproduce the original signal.

This chapter begins with a discussion about digital video, followed by an overview of its historical development over the years. Different digital video standards and consumer devices offered by the major computer vendors are examined to illustrate what is available. Apple Computer's solution to digital video, QuickTime, is discussed briefly to see how it overcomes and manages the issues of controlling time-based information. Alternative digital video formats are examined to find out how they manage time-based information. The different technologies are compared and evaluated to find out if and how they could be used to assist in the analysis of time-based information. Their strong points will be discussed along with any major disadvantages. Human-Computer Interface issues of the different technologies are also discussed.

3.1 What is Digital Video?

Digital video is the means of storing video images and audio onto digital based equipment such as computers. Television broadcasts and video tape information at present are both examples of analog technology. In our computer based society the majority of this information cannot be incorporated into our everyday transactions with the computer.

Analog video signals have to be converted to digital signals in order for the computer to use them. The digital computer signals have to be converted into analog signals to display them on a video monitor or for recording. For digital video to be successful, the computer must be able to convert analog video signals into digital data in real-time.

As an indication of the amount of information needed to be processed by a computer, a single frame of video data can easily take up 1 MB of disk space when converted into digital data. Television displays 25 frames[1] of information per second. If this was to be stored on a computer then it would require approximately 25 MB of information for every one second of video recorded without any audio being considered. This implies that about 25 MB of data would have to be transferred from a storage device and displayed every second. This is faster than any desktop computer can handle. Internally, a Macintosh computer can only transfer around 10 MB[2] of data per second to some hardware components while an average hard drive transfers around 1.5 MB of data per second. In comparison to these figures, most Local Area Networks only operate at 10 MB per second. It is obvious that computers in their current form are unable to handle what is required for this sort of data transfer in its raw form.

Even if the computer could manage this sort of data transfer, the sheer size of the data would fill up even the largest of hard drives in a very short time. To store one minute of video data in its natural form would require a 1.5 Gigabyte hard drive which is nearing the limit in storage devices commonly available at the present.

There are multiple ways to overcome this major problem. The simple method is to reduce the amount of information needed to be processed. The reduction in the number of frames that have to be stored per second can significantly decrease the amount of information to be processed. Reducing the physical size of each image on the screen and the number of colours of each image both reduce the amount of information. Even with all of these techniques employed, there is no guarantee the amount of information is reduced enough to be successfully processed. A more recent technique that is commonly employed in digital video is the use of compression techniques in real-time. This technique will be discussed in detail in the following sections after examining the history of digital video.

1 24.99 frames per second is the PAL (Phase Alternate Line) standard which is used for New Zealand and English television. In comparison, 30 frames per second is the NTSC (National Television Standards Committee) standard as used in the United States and Japan.

2 10MB is the data transfer rate that NuBus cards operate at. Video cards and digitising cards are examples of this type of card.

3.2 A Brief History of Digital Video

> *"Multimedia (or 'many media') is a category for a whole gamut of tools, ideas and skills that can be used to convey or access information."*
>
> \- Tony Crease [5]

The word 'multimedia' is a term much used and abused by the popular press, creating a great deal of confusion amongst the general audience of computer users and specialists alike. It is a general term, used to represent an environment in which more than one media-type is used, media-types being text, graphics, animation, audio and video. Video tape is 'multimedia' since it can deliver text, graphics, animation, audio and video. The key difference we have with the computer-based environment is that all these media-types have been digitised, allowing for instant random-access to information that may be represented in any, or a combination, of these media types. This is known as media-integration, where one is able to integrate and control all of these media types on the desktop. Multimedia is not new, it has been with us in various forms for many years. The marriage of images and sound in television and film, for example, are simple examples of multimedia.

Text, pictures and sound on computers have been a common feature for many years now. The next step in this evolution is to have moving images. Many simple proprietary schemes loaded a series of images into memory and cycled through them. Once the first sequence had been played, the next sequence was loaded and then these were cycled through. This process gives the user the impression that the video is playing back in real time. However this is not the case. Real time playback involves the ability to continuously drag the information directly from the storage media and display it without any delay.

In the past, the lack of industry accepted standards for digital video led to many proprietary schemes being used. Subsequently, many of these did not survive long in the market place due to the schemes not being accepted by the large consumer market. In order for digital video to survive and develop further, standards had to be developed and agreed upon. Once this was achieved more applications could be developed in the security that they would last in the environment being developed for. More features for the hardware and software could then be developed and consistent interface issues could be addressed so that the handling of digital video data could become as easy and familiar as handling text and graphics is now.

The evolution of full-motion digital video was not instantaneous. It was a gradual progression from the capture of static images leading through progressively more improved technological stages until reaching the standards that we are now familiar with.

Video digitisers were the first type of product which integrated the video and the computer. This type of product converts analog video into digital form. This process usually required between 10 and 20 seconds to complete the digitising of a single frame. The results of this was a static image which was only useful for scanned documents and illustrations. It was not feasible to capture moving images due to the time required for each frame.

The next stage in the evolution of full-motion digital video was the frame grabber. This essentially gave the capability of grabbing a single frame in real time of the 25–30 frames that pass by every second. This process worked in memory and the only limitation was the amount of memory available to perform this type of operation. If no free memory is available then the frame grabber will not work. Since the digital image is being stored into RAM, the frame grabber cannot capture video continuously. The images captured are again only useful for illustrations and presentations. Richard Olivo [33] used a black and white frame grabber in the analysis of animal's movement using this technique.

Another technological improvement was video-in-a-window. This method displays analog video signals into a window on the computer in real time. The analog video signals are converted into the analog RGB display format of the computer monitor and integrated with the monitor output at the display-card level. The main problem with this method is that the signal is not converted into digital form but simply passed directly to the display-card. Therefore the computer cannot access any of the video information. The only real use for this type of technology is for watching the source of the analog video signal. Video-in-a-window could be used to display the output of a video tape being played back onto the computer screen while using the computer to either control the video tape player or using the computer to assist in taking notes.

Since there was a lack of digital video standards in the early pioneering days of this technology, the industry utilised existing graphics formats which had been developed earlier for static images. This seemed to work well as video digitisers and frame grabbers worked with these still images. These formats can only represent still images so their value is relatively limited. A few proprietary formats had been developed but they only simulated full-motion video. The formats were secret and the user could not manipulate the video contents. These formats were not used since full integration was impossible.

Digital image compression is a viable solution for effectively reducing the file size of large still images. This provides a way of storing large numbers of images onto a standard hard drive which would previously be impossible. Not only is the physical disk space reduced but the time required to display and manipulate the images is reduced since their sizes are smaller.

The most widely accepted standard adopted for digital image compression is JPEG [36]. JPEG based compression can significantly reduce the size of the image (as high as 24:1) while still preserving the image quality. With a compression ratio of 20:1, the previous 1 MB video frame would be reduced to a mere 50 KB. This would reduce a second of video data from 25 MB down to 1.25 MB. A minute of video data would be reduced from 1.5 Gigabyte down to 75 MB of information. Even though this is still a significant amount of data, the results are achievable with today's technology. With the use of compression, a series of frames can be processed and displayed by the computer quickly enough for the movement to appear realistic. This is the basis of motion JPEG (MJPEG) where each frame is compressed using JPEG compression [38].

With current storage devices, it is impossible to read the amount of information required for full-motion video. A typical fast hard drive can only retrieve about 1 megabyte of information per second. A standard CD-ROM player has a through-put of 150 KBytes per second. Standard storage devices are not able to cope with the requirements of digital video. Even if a drive was developed to handle this, or multiple drives were connected together in a parallel fashion capable of this data rate, there is still the limitation of the protocols for disk transfer. SCSI for instance has limitations on the amount of data that it is capable of handling. The best way to store this amount of information is by compression. This compression must be able to reduce the size of information so that it can be effectively retrieved from CD-ROM at 150 KBytes/sec and it must also be able to decompress the information in real-time.

Typical compression methods employ both spatial compression, the compression of an individual video frame, and temporal compression, storing mainly just the information that changes between subsequent frames. These methods are now the basis of modern digital video techniques. More information about spatial and temporal compression can be found in Chapter 4 on *QuickTime*.

Other ways to reduce the amount of information include reducing the size of the image, the image quality, and the number of frames to be displayed per second. These all contribute to reducing the amount of information to be stored and retrieved in real time.

Many schemes require additional hardware to assist in performing the compression and decompression. QuickTime, Apple's response to digital video, does not require any additional hardware to perform compression and decompression of digital video information.

QuickTime is a free system extension to the Macintosh operating system. Its purpose is to provide the integration of digital video and sound into the standard work environment of users. QuickTime works behind the scenes and is transparent to the user. Digital video elements can be integrated into applications and documents just as text and still image graphics have been for many years using the copy and paste metaphors.

QuickTime has been chosen as the framework for this thesis for several reasons. QuickTime is free. All other systems require the purchase of additional software or hardware to perform digital video playback. The format is the standard on the Macintosh and is available on many other platforms. QuickTime provides the compression and decompression of motion video transparent to the user. It is entirely software based but is capable of taking advantage of any hardware that can assist in compressing and decompressing video information. Users can place the system extension into their system folder on their machine, restart their machine, and immediately take advantage of digital video. The last important factor in the decision of using QuickTime is the platform that it runs on. Macintosh computers are commonly available and more importantly were available in the Psychology and Computer Science Departments that were involved in this research.

A more in-depth analysis of the structure and functionality of QuickTime is available in chapter 4.

QuickTime is but one of the many standards currently available for digital video. In recent years several of the major computer companies have introduced their own implementations on various platforms. These major standards and implementations will be discussed briefly.

Not all of these offerings from the major computer companies are directly comparable with QuickTime but all show a form of digital video that is worth mentioning. Some offer more features, others are more expensive, and some are not appropriate in assisting in the analysis of time-based information. The support from the major computer companies of these schemes is examined with a look at their incorporation into existing platforms and the future as far as becoming well accepted standards. These following standards are categorised by their associated company names.

3.3 Microsoft, Inc

Microsoft, Inc is one of the largest computer software companies with a major influence on how the computer industry operates. Microsoft's main direction in the digital video market is in establishing a niche in the market for its Window's Multimedia Extension (WME) and Video for Windows products. WME is an extension to its already very popular Windows product that has a large foundation of users in the IBM compatible computer market. WME provides the needed interface between multimedia applications such as Video for Windows and Microsoft's Windows environment. Video for Windows is Microsoft's product that provides digital video to the IBM-PC compatible market. Along with the introduction of WME and Video for Windows, Microsoft is trying to establish the Multimedia Personal Computer (MPC) as the standard delivery platform for multimedia content. Together this hardware and software combination represent Microsoft's platform for delivering digital video to the PC market.

Multimedia — and digital video for that matter — need more than a basic PC. This minimum configuration is the MPC which is essentially a normal PC with a CD-ROM unit and the Windows Multimedia Extensions.

Microsoft's approach is for the user to buy special hardware and software in order to run multimedia based software. Microsoft believes that the standard PC is for "serious" work. If one wants to run multimedia based content then the user has to go out and buy a special computer.

Since the market is more demanding of "standards", Microsoft's approach with the MPC has been welcomed by many customers, who believe that Microsoft will be successful in creating MPC as an industry standard. Many people, in the PC world, believe that WME is the standard for multimedia technology in the PC world. However no true hardware and software standards exist in the PC world which subsequently causes confusion and difficulty for the consumer and developers. Microsoft themselves believe that MPC is the ultimate standard yet other major companies like IBM, Compaq and Dell are not supporting MPC. The same can be said about future operating systems on the PC. There is no clear answer as to which one people should be using. There are many to choose from, including OS/2, Windows, Windows NT, MS-DOS and several others.

MPC supports little integration of hardware and software since no-one is in control of both areas. Many vendors who provide hardware and software are not even involved with MPC itself.

There are many barriers to MPC becoming totally accepted. The major barrier is its cost. An MPC computer costs around $US800 more than a normal PC [2]. This extra cost for users may be an added barrier.

Microsoft sees multimedia as a niche in the market. Their strategy with MPC is to "bolt on" existing technologies and call it multimedia. There is little media integration into general productivity applications. For example a user cannot copy and paste time based information to another application like a word processor. MPC and WME are an attempt to bring the IBM PC up to the level of the Macintosh circa 1988. Apple's technology, even without the addition of QuickTime, is capable of what MPC provides. In addition, a Macintosh with QuickTime will offer not only simple playback of digital video, but also integration of these media types through **all** general purpose applications. Currently there is no concept of a time based architecture for WME that even resembles QuickTime.

Table 3.2 is a comparison of technologies between Macintosh and the Multimedia PC.

Macintosh 1988	**Multimedia PC 1991** DOS, Windows, Windows Multimedia Extensions and MPC Hardware
Display Hardware - 8 Bit Graphics (256 Colours at 640x480) easy to upgrade to 24 bit	VGA - 4 Bit Graphics (16 Colours at 640 x 480) difficult for users to migrate up to 24 bit without re-configuring and adjusting the software and hardware settings.
Display Software-Colour QuickDraw followed by 32 Bit QuickDraw (1989) Transparently provides a standard auto configuration for various monitor depths and sizes and also provides for extremely fast bit mapped graphics display.	Many different third party monitor drivers in addition to palette editors that the user must install and configure depending on the Hardware used. Plethora of graphics standards (EGA, CGA, VGA, VGA+, 8514A, XGA) without graphics architecture.
Stereo Sound Playback and Sound Manager	Third party cards such as the Sound Blaster Card from Creative Labs, Inc.
Sound Digitising (MacRecorder™)	Sound Blaster Card w/ additional Microphone
HyperCard CD-ROM Toolkit	Media Control Interface for CD
HyperCard Videodisc Toolkit	Media Control Interface for Videodiscs
Apple MIDI Manager	Media Control Interface for MIDI

Table 3.2. Comparison of Macintosh Technology (1988) with Multimedia PC (1991) from [2].

Table 3.2 is a comparison of a Macintosh computer in 1988 with a Multimedia PC from 1991. We can see that the Macintosh offered a more integrated and easier to use environment at a much earlier time than the Multimedia PC had to offer three years later. The Multimedia PC required the purchase of multiple hardware components to offer the same capabilities of a similarly powered Macintosh.

Macintosh 1991 with QuickTime	**Multimedia PC 1991** DOS, Windows, Windows Multimedia Extensions and MPC Hardware
JPEG Still image compression	
Animation Compression	Linear playback of "Director" animations
Software Video Compression	
Open Architecture for additional hardware and software image compression schemes through Component Manager	
Open Architecture for video digitiser cards through Component Manager	
Open Architecture for ANY other system addition (e.g. printers, colour-matching schemes, storage devices, system software extensions) through Component Manager	
Time based architecture which provides synchronisation throughout the system. Key to media integration	
Technologies are part of Standard operating system, not a subset with a niche market.	

Table 3.3. A Comparison of Macintosh With QuickTime and the Multimedia PC in 1991 (From [2]).

Table 3.3 shows the advantages that QuickTime has over the Multimedia PC. The Multimedia PC does not offer a time based architecture or the total media integration that the Macintosh offers.

To upgrade a PC system to effectively become MPC compatible requires the WME and a CD-ROM drive. In comparison all that is required on a Macintosh is to drop QuickTime into the extensions folder. QuickTime itself is free and no additional hardware is needed to operate multimedia based software.

Pricing of the two technologies is approximately the same once all the additional hardware has been purchased for the Multimedia PC. Both platforms have seen dramatic reductions in their prices over the last several years as the major computer vendors compete in their price wars.

In late 1992, Microsoft announced a new technology for dealing with time-dependent media and digital video for the Windows system, 'Video for Windows', which was previously known as 'Audio Video Interface' (AVI). This provided a scalable video capture and playback architecture allowing the capture of video in a non-compressed format. From there it could be edited and saved into Microsoft's compressed AVI format. To use Video for Windows, the user must have Windows 3.1 and the WME. WME is the multimedia software tools needed by Windows 3.1 for Video for Windows to work. It allows playback performance slightly worse than QuickTime 1.5 on a 16MHz 386 MPC style computer. For capturing video, a faster (33 MHz 386) computer is required together with a third-party video digitising card using the Intel i750 video processor [4].

3.4 IBM

IBM's main stake in the digital video market is through their Audio Visual Kernel (AVK) system software used in conjunction with Intel's Action Media II board that uses the Digital Video Interactive (DVI) compression device [38].

The Audio Visual Kernel is system software that enables digital video under Windows and OS/2. It is a multi-layered architecture that isolates hardware-specific features from the application programmer while enabling porting of audio and video to other platforms. It provides a low-level programming interface that is designed to integrate smoothly with multimedia programming interfaces such as IBM and Microsoft's Media Control Interface (MCI) and Apple's QuickTime. AVK is sandwiched between Window's API and the Action Media II hardware that contains the i750 video processor [7, 9, 17].

Digital Video Interactive (DVI) is a video compression technology originally developed by RCA [21], now owned and being developed by Intel. Digital Video Interactive consists of hardware based on Intel's i750 video processor. This processor allows the playback of full-screen, full-motion video. It is expensive and the compression and decompression is not fully integrated, requiring the purchase of an additional delivery and capture board that fits into a slot on the DVI board.

The i750 video processor consists of two processors; the 82750PB, a pixel processor, and the 82750DB, the display processor. The processor is controlled by a micro code engine which is a collection of routines that are loaded into the instruction memory aboard the pixel processor. These routines consist of the compression and decompression, scaling, and copying routines.

The digitised video and audio data are compressed using different algorithms. There are three types of compression for video images available, Indeo [38], Real-Time Video (RTV), and Productions-Level Video (PLV). The Real-Time Video (RTV 2.0) algorithm provides compression at a resolution of 128x240 pixels by 30 frames per second (fps) (NTSC) or 128x244 pixels by 25 fps (PAL). RTV doubles the number of horizontal pixels on playback resulting in a 256x240 (NTSC) or 256x288 (PAL) video window. RTV is implemented in micro code and executed on the pixel processor of the i750. It is a lossy algorithm that makes multiple passes over the video data using several techniques including frame-differencing and Huffman coding to reduce the bit rate. RTV 2.0 improves over the original algorithm providing better image quality, and adjustable data rates and image quality [14]

PLV uses essentially the same compression as RTV, but uses off-line compression services to gain similar speed performances. PLV can achieve performance similar to that of RTV, [9]. Audio data is compressed using a 4-bit adaptive compression algorithm (ADPCM4). This scheme effectively works by predicting the next sample based on the previous sample [9].

The most difficult part of capturing the incoming digitised data for AVK is keeping up with it. If it does not read the frames from the VRAM buffers fast enough, the frames are lost, and a series of blank frames have to be inserted to take their place in order to keep the frame rate constant. This process causes skipping effects on the playback of the video data.

DVI supports three digital video formats: QuickTime, AVI (Audio Video Interface), or more commonly known as Video for Windows), and AVS (Audio Video Standard).

DVI and AVK are still a "bolt on" approach offering no integration with the rest of the system. IBM's rift with Microsoft has lead to confusion about who to turn to for standards. IBM often supports "all standards", thus they do not have any concrete direction or vision. IBM themselves have had a difficult time making OS/2 the standard operating system of the future and their digital video system is only one of many available [14].

With IBM and Microsoft being major competitors in the PC based digital video market there is a strong feeling of uncertainty amongst developers and customers as to what the standard is. Apple, IBM, and Motorola have signed a letter of intent for the development of industry standards for multimedia. The involvement of these three major companies is of no surprise since their involvement in the development of the new PowerPC computers to be released in March 1994.

QuickTime on the Mac and AVK on the PC both provide time-oriented, data-handling operating-system extensions. There are similarities between the two systems. For example a comparison can be made between the main features of QuickTime and AVK.

QuickTime's Component Manager, which allows a runtime binding of code modules into applications is similar to dynamic linked libraries under Windows. The Image Compression is similar to AVK's MVD. Finally, the third major feature of QuickTime's structure, the Movie Toolbox, which is the primary application program interface (API) that contains all the calls needed to play and record dynamic media, is similar to Window's Media Control Interface (MCI). The details of QuickTime can be found in Chapter 4 — *QuickTime*.

AVK is based around a system requiring hardware assistance, namely the i750 video processor. QuickTime, however is designed as a general purpose multimedia extension of the Macintosh operating system. It does not presuppose any hardware but can take advantage of any that become available. Compression is usually achieved by one of the many CODEC (compression/decompression) components normally done entirely in software. AVK's compression algorithms, RTV and PLV, are implemented in i750 micro-code, which is difficult to modify [10].

3.5 SUN Microsystems

By end of 1992 one million SPARC stations had been sold. Though this may seem an insignificant amount of machines compared with of the PC and Macintosh market, it is significant in the digital video market. SUN Microsystems have their own multimedia integration group with their focus on networked multimedia solutions for video. It is rumoured that they have a number of software only video compression and decompression schemes under development. Larger video window sizes and faster playback than on Macintoshes and PC's is achievable due to the fast processors and system throughput of these RISC based workstations.

Of course, with the increased speed comes the increased price to match it. SPARC stations come with 8-bit graphics as standard but this is moving towards 24-bit video, using the Intel i750 chip set standard on the motherboard as general purpose video processors.

This type of hardware platform is better suited for video and animation processing due to the fast hardware available, larger monitor sizes, and the movement towards 24-bit accelerated video as standard. Other important features that are also standard on the SPARC stations include ethernet for networked multimedia, and sound processing hardware.

A major drawback for SUN is its poor distribution in terms of mass market consumer items. The products available are not general for consumer usage, but only specialised multimedia applications are available. UNIX itself, even with Open Look or similar graphical user interfaces available, is not user-friendly for an average user.

With all the extra power from these machines, the price is proportionally higher. The lowest cost SPARC station is $US4,995 which is $US3,000 more than a Macintosh LC, making them an expensive machine for the average digital video consumer. SPARC stations may be fast by today's standards but hardware speed and performance is increasing all the time. With Apple, IBM, and Motorola's joint venture, the new RISC based PowerPC machines being planned for release in March of 1994, aimed at the consumer market, offer an enormous amount of power for a fraction of the price of a SPARC station [2].

3.6 Philips, Inc

Philips, the creators of the CD audio standard have been working on Compact Disk Interactive (CD-I), a consumer based electronics multimedia standard.

The whole approach to CD-I is from a consumer product orientation. Philips are trying to create an industry specification like CD audio that can be scaled to different platforms such as TV's or PC's and could be licensed and used by anyone. They have been promoting the CD-I standard since 1986.

Delayed shipments and development of the Compact Disc Interactive (CD-I) left the financial stability of Philips in question. They have been trying to establish CD-I as an industry accepted consumer electronics standard for interactive multimedia. The development of CD-I titles is an expensive process costing anywhere between

$US9,000 and $US25,000 [2]. They are also time consuming and difficult to create. The CD-I standard is also supposed to be able to support multiple vendor's products like Kodak's photo CD and Nintendo games.

Philips are working on the CD-I in conjunction with Sony and Matsushita even though these companies are working on competing products like the CDTV and DataDiscman.

One of the biggest selling points for CD-I was to extend the standard to allow full-motion video through the addition of video decoding hardware [41]. This was achieved through hardware capable of decompressing MPEG [12] files, a standard for compressing full motion video. Since there were delays in the standard being defined, there were delays in the production of the CD-I. Major drawbacks to the CD-I standard stemmed from the delays in the shipping of the MPEG component of it. This meant that MPEG could not be included initially and relied on an expansion slot for a card upgrade to existing systems.

There are many standards available like CDTV, CD-I, Sony's DataDiscman. It is unclear — and unlikely — that any one will become the ultimate standard. When CD audio was introduced there was no competition. Now that Philips are trying to promote their CD-I standard they have some major competition. Sony, one of Philips major backers in the CD-I standard, has its own product that shows a lack of confidence in Philips. Matsushita, the other major backer in the CD-I standard also produced CDTV with Commodore.

At the time of the release of CD-I there were only 50 titles available for consumers to buy. To appeal to the mass market there needs to be 1,000 titles available. The development costs of titles makes it prohibitive for small publishers to produce their own titles. Philips had been pre-announcing the completion dates of the CD-I since the introduction in 1986, none of them had been met. This has led Philips to have a lack of credibility, and companies turned to other platforms while they were waiting for the big release of CD-I [2].

3.7 Commodore Business Machines Ltd

Commodore's contribution to the digital video market has been the CDTV, and recently, its successor, the CD^{32}. CDTV was a repackaged Commodore Amiga 500 with a CD-ROM, connected to a TV as a consumer device. The long term commitment may be difficult since Commodore is not the most financially successful company. It has a poor reputation due to its lack of huge commercial success with the Amiga product family, although the C-64 did quite well in the home market.

CDTV brought information and entertainment to life, enabling the viewer to do, see, and visit things in a way that couldn't be done before. Another benefit of the CDTV was the fact that it could also be used as an Amiga 500 computer. This was a benefit over the CD-I which can only be upgraded to a limited PC.

The size of the CDTV hardware was about the size of a standard CD player, with an infra-red remote control connected to a television set. The price of the hardware is $US950 and individual titles range from $US25 to $US75 each. Unlike CD-I, the CDTV uses software based video decompression but at a lower quality than QuickTime 1.0. CDTV was not a technical breakthrough or a revolution in consumer electronics that offers new features. It was simply an Amiga 500 computer with a CD-ROM unit attached. It was based on a machine that was introduced in the second half of 1987 for about $NZ2,000, [2].

The CDTV was not a success for Commodore with very poor sales. It was discontinued but has recently been replaced by the CD^{32}. The CD^{32} is an Amiga computer just like the CDTV. Its main purpose is as a consumer device, being marketed as a games console to compete with the major companies producing games machines like Nintendo and Sega. The CD^{32} also supports many other standards like the earlier CDTV format, CD+G (Karaoke type graphics), audio CDs, Philips' full-motion CDs, and CD^{32} titles. It delivers full-screen, full-motion video playback with the purchase of an MPEG plug-in module. Essentially the CD^{32} is only a playback device with some interactive capabilities. It is priced at $NZ899 without the MPEG module which is somewhere between a cheap games machine for home use and a home computer. Like the CDTV, the CD^{32} has the capability to be used as a computer.

3.8 Other Standards

With the advent of CD-ROM technology, there has been a number of formats arise. Several of these have already been discussed like those used in the CDTV, CD^{32}, and CD-I. Others include CDXL, CD Video, Kodak's Photo CD for still images and others. Many of these formats have been defined due to the lack of one universal format that is freely available for everyone to use and is capable of everything that may be required. Most of the formats that have evolved for CD-ROM digital video technology have been mainly used for video playback like those for laser disks. There have been several formats that have included facilities for an interactive environment. Such formats may be used in interactive digital video games whereby the player can make choices as they play the game.

Many of the technologies will support multiple formats so that they are more competitive in the marketplace and are not left behind by being incompatible.

3.9 Costs and Comparisons of Technology and Features

From Section 3.4 we see that there are many different digital video formats available from the major computer vendors. Many of these schemes are dependent on specific platforms and are not portable. Whenever a format becomes large enough in their own right then other companies usually have to support it to be competitive in the marketplace. There is the possibility of bigger sales for the other companies if they can say that they can support multiple formats if their competitors cannot. For digital video to become truly accepted by all computer companies, there needs to be a format that is easily transported to different machine architectures and independent of any particular hardware. The resulting digital video files must be interchangeable amongst all platforms.

The cost of hardware and software required for digital video is a major issue. To get people to adopt standards widely, they should be easily affordable to almost everyone, with as little setting up as is possible for the user.

Many of the systems examined offer a common set of features. The most common is the playback of digital video. There are strengths and weaknesses of each system. Some offer very little in features except video playback while others offer totally integrated environments with the rest of the operating system making the digital video enhancements transparent to the system and the user.

3.9.1 System Comparisons

Comparing systems with each other is a difficult task. Each system generally has many unique features that the others do not offer. Some features, like the cost of hardware and features offered, are easily examined. The computer based systems like those offered by Microsoft, IBM, and Apple all have similar costs for the computer hardware. However, Microsoft and IBM require additional hardware and software for digital video capabilities whereas Apple's QuickTime is a software based solution.

SUN Microsystems also offer some highly specialised digital video solutions but the hardware requirements are very high.

The CD-ROM based solutions vary in price and performance. Philip's CD-I suffers dramatically in the area of the cost of producing titles. Commodore's CDTV and CD^{32} are both reasonably priced platforms with the creation of titles at an attainable cost. Both can create the titles by using regular Amiga computers to generate the material for the CD-ROM titles. The CDTV and CD^{32} are essentially interactive playback devices.

3.9.2 Media Costs

Digital video offers many benefits that other analysis methods are unable to offer. Advantages include the ability to link the analytical transcriptions with the actual video information, instant access to the video information, and analysis of multiple video windows to mention just a few. The major disadvantage of using digital video is the disk space required to store the digital video information.

Digital video can be a relatively expensive solution compared to traditional video tape analysis methods. The cost of the storage media is a large factor to consider when comparing the advantages of digital video over other traditional methods. Normal digital video media would consist of a hard drive or magneto-optical storage devices to hold the information, while video tape is the preferred storage media for analog systems. Digital video is limited to expensive types of media. Alternative storage media include CD-ROM, and digital tape media like DAT. These alternative media are viable storage methods and deserve some careful consideration.

A typical average quality VHS video tape costs around $10 in New Zealand.[3] This holds three hours of time based information that can easily be used on any video player for analysis work. This works out at a cost of 5.6 cents per minute of video information. This will produce full screen images at 25 frames per second[4], the quality of the image is dependent on the video player that the information was recorded on. Video players are the typical equipment that a large number of researchers use in their analysis of time based information.

3 TDK 180 minute VHS, PAL video tape.

4 Phase Alternating Line (PAL) standard.

The cost of storing digital video is considerably higher. To make a comparison we will assume that one minute of digital video information is 5 Megabytes in size. So to hold 180 minutes of information requires a 900 Megabyte storage device. The price of a hard drive to hold this would cost approximately NZ$2,000. This implies that one minute of information costs approximately NZ$11.11. This is about 200 times the cost of video tape media to store the same amount of information. This high cost makes digital video seem an unattractive option. The benefits of digital video have to be considered carefully when considering a system with such high media costs as to whether the advantages that digital video can offer out-weigh the cost of the high media.

The benefits that digital video offer include the ability to access any location of the information instantly; having multiple video and audio components that can be viewed simultaneously, which enables better analysis techniques; and audio analysis capabilities, video analysis capabilities such as zooming, and full control of playback direction and speed.

The cost is not the only limitation of digital video. The image size is normally about a quarter of the screen and is not broadcast quality. This is not necessarily a major problem since the nature of the material does not always require such high quality images. For example, in analysing children's behaviour, the observer may only be interested in the child's movements with the other children which does not require extremely high detail of images.

Depending on the nature of the research, the cost may be easily justified. However, not everyone has this much disk space available all of the time. There are alternatives to using a hard drive as the main source for examining digital video. Two alternatives are CD-ROM and digital tape based media.

The creation of 'one-off' CD-ROM disks is now a feasible and inexpensive way for storing digital video information. It is now becoming a very popular way to store large volumes of information at a very minimal cost without losing instant access to the information that is associated with tape backup methods. At around NZ$37 for the blank CD-ROM disk, this option is a very realistic approach to storing digital video at low cost. Each disk can hold approximately 650 Megabytes of information. This translates into approximately 130 minutes of video information, costing about 21 cents per minute of digital information. Once a CD-ROM has been successfully produced, the original data can be erased from the disk drive, freeing up a valuable and expensive resource. Speed of access is slightly slower than a hard drive but it is nowhere near the speed of searching through a video tape. One disadvantage is that the media cannot be re-used. If the information is to be kept anyway then it is not an issue. CD-ROM disks have a longer life span than video tape and do not suffer from wear or deterioration, nor can they be accidentally erased. For digital video to become a realistic option, CD-ROM is the most cost effective method available. This assumes that access to a CD-ROM mastering device is available. This is starting to become very popular at the current time.

Another possible approach is to store digital video is the use of digital tape backup units such as DAT or Exabyte formats. The digital video can be read and written directly with the device almost acting as a video tape player. This method will lose some of the advantages that the other digital video media can offer like instant access to information. It does however offer other capabilities of digital video like multiple video and audio components that can be analysed simultaneously. Audio and video analysis are also available together with playback controls. Another important factor is the cost of the tape itself. A typical DAT tape drive will hold about 2 Gigabytes of information. This works out to be about 400 minutes of information that can be stored on a single tape. A single tape costs around NZ$35, or 8.75 cents per minute of digital video information. This is comparable with that of normal video tape media which is only 5.56 cents per minute.

Table 3.4 shows a summary of the different media costs discussed.

Media Type	Media Cost ($ / Minute)
Video Tape	$ 0.06
Hard Drive	$ 11.11
CD-ROM	$ 0.21
DAT Tape	$ 0.09

Table 3.4. A comparison of media costs.

From this we can see that tape media is easily the cheapest form of storage for both conventional and digital video based analysis methods. The cost of using a hard drive is very high. Prices for hard drives are dropping each year for storage devices. The actual cost for a hard drive is not excessive but as a comparison with alternative storage media it is very high. Based on trends over the last several years, the price of hard drives will continue to decrease and increase in size for the low-end models. Hard drive requirements by consumers has been steadily increasing due to software getting larger and users starting to explore static and moving digital data.

Another comparison to consider is the cost of the hardware devices that use these media. In the area of video players, two categories are available; standard home use type and computer-controlled players. A standard video player for home use can cost as little as NZ$500 while a sophisticated computer controlled video player can cost up to NZ$4,000.

A major consideration for many of these systems is whether they require a computer. The analysis method that uses a standard video player may not require a computer, but many researchers still employ the use of computers to assist in their analysis of their results, so the purchase of a computer could almost be viewed as independent of the method chosen. New computers range from NZ$2,000 and upwards for a basic personal computer such as a Macintosh or an IBM compatible PC.

Schemes that are based around digital video need a video digitising card to convert the analog video signals into digital format that the computer can understand. For a basic video digitising card the prices start at around NZ$1,000. It is not always necessary to buy a digitising card. It is quite likely that one could be borrowed or rented since the digitising stage does not take long and may only be performed infrequently. The majority of the time is spent analysing what has been digitised.

A hard drive capable of holding 180 minutes of digital video may cost about NZ$2,000. Holding 180 minutes of digital video information is an expensive exercise. Another consideration is what to do with the information once it is not needed. Should it be backed up for future reference or removed? A tape backup unit may be used to store this information or even make a couple of CD-ROM disks out of the information.

The hardware required to master single CD-ROM disks can cost around NZ$7,000. This is an expensive option for an individual but access to this type of machine is becoming very widespread in New Zealand. With access to a mastering machine, the creation of an alternative random access media is relatively cheap. This can replace the need for a large hard drive, although one is useful to make the original source information on before pressing it to CD-ROM disk.

Tape backup units cost in the region of $NZ2,500 for a DAT backup unit. The tape unit can be used in one of two ways: either as a backup of the information on a hard drive or as a primary source of the digital video information. By backing up the contents of the hard drive when the source digital video information is finished with, the researcher can restore it onto the hard drive at any time. In this way the hard drive acts as a scratch disk which is only used as a temporary storage medium. As mentioned previously, the tape unit can also be used directly but cannot provide fast random access.

3.10 Human-Computer Interface Issues

> *"Consumer companies are more concerned with production and manufacturing at low costs and have little concept of human interface. For example millions of people to this day still do not know how to set the clock on a VCR. Apple's forte is in software development such as hypercard and Human-Computer Interface design. These two are crucial to the success of multimedia."*
>
> - Don Crabb [4].

The majority of digital video systems are essentially based on the controls found on video tape players. These features supply such functions as play, stop, fast forward, and rewind. The actual interface design of these elements differ significantly from platform to platform and also the applications on each platform. There is no consistent interface for users of digital video. QuickTime itself does not escape from this observation. The many products available on the Macintosh all seem to implement their own style of controls for digital video. It almost resembles the status that independent flavours of digital video implementations suffered before the introduction of standards. There were many implementations of digital video available but all were proprietary schemes and were incompatible with each other. The introduction of standards to digital video has alleviated many of these problems, even though there still is a large number of digital video standards. However the next step in this is to implement consistent controls for manipulating digital video data.

3.11 Summary

QuickTime offers the most suitable features as a platform for further development in the analysis of time based analysis methods. The decision to use QuickTime for this project was based on many factors.

Some of the factors for deciding on using QuickTime included the resources for development work available; namely Apple Macintosh computers. The freely available technology, QuickTime, was also another consideration in deciding the platform. The only cost involved was the purchase of the development kits for

QuickTime. The freely available technology is a real advantage for development work, users of the technology do not have to purchase additional technology in order to use the software written. The expense of other alternative systems was a discouragement when equal or better alternatives are available with QuickTime.

Chapter 4

QuickTime

QuickTime is Apple Computer's response to presenting time based information on the desktop, an extension to the Macintosh operating system. The introduction of QuickTime in December 1991 offered an ad-hoc community of digital video users a standard video digitising interface. The nature of QuickTime allowed video digitiser manufacturers to concentrate on making value-added hardware and software, secure in the knowledge that their products would work with whatever general purpose video capture and editing applications were on the market. Application writers could at last code to a standard interface and take advantage of improvements in the underlying hardware as they came along. Customers also benefited from a standard that usually simplified interaction with these devices.

The QuickTime system software extension adds capabilities that let application programs integrate graphics, sound, video, and animation into documents. By providing a standard way for all Macintosh programs to control these multimedia elements, QuickTime makes them easier to use.

Users have no direct interaction with the QuickTime system extension. It works behind the scenes as part of the system software. Rather, it provides new capabilities to the application programs that support QuickTime, transparent to the user.

This chapter examines what QuickTime offers to the world of digital video. Initially the structure and functionality of QuickTime are examined, followed by a look at the Human-Computer Interface offered. Following this is a history of QuickTime's release and subsequent updates. The availability of QuickTime on other platforms is also examined with a look at the future of QuickTime and where it is leading the desktop of Macintosh users.

4.1 What is QuickTime?

QuickTime provides a system-level architecture for manipulating and synchronising full-motion images and sound. Video is now an integral part of the Macintosh operating system. QuickTime is essentially a media-integration architecture with provisions for video, audio, animation and device control.

A traditional movie, whether stored on film, laser disk or tape, is a continuous stream of data. A QuickTime movie is constructed similarly. The term *movie* in QuickTime is not referring to the medium; it is the organising principle of time based data. Similar to traditional movies, a QuickTime *movie* may contain several *tracks.* Just as a cinematic movie can contain several tracks (a visual track and a sound track), a single QuickTime movie can contain more than one stream of data.

QuickTime is an extension of the Macintosh system software that enables the integration of time based data into mainstream Macintosh applications. Examples of this type of data include video, sound, animations, data produced by scientific instruments, and financial records. In QuickTime this set of time based data is referred to as a movie. To understand more about how QuickTime handles time, time bases need to be examined. Time management is an essential part of QuickTime. The time co-ordinate system anchors movies and the media data structures to a common temporal reality, the second. Each time co-ordinate system contains a time scale, marked in time units, that provide the translation between the time in a movie. The number of time units that pass per second quantifies the scale of the movie. For example, a time scale of 25 means that 25 units pass per second and each time unit is 1/25th of a second. The time co-ordinate system also contains the duration of the movie represented by the number of time units. All of this information is found in the header information of each movie.

The time base can be viewed as a "black box" that maintains the temporal characteristics of a QuickTime movie. These temporal characteristics consist of the movie's duration, whether it is moving or not, whether the movie is looping, and the clock that controls the time. The time base for a movie is created dynamically when a movie is opened.

Most movies have a time scale of 600, meaning that the smallest fraction of time measurement for the movie is 1/600th of a second.

The *rate* of a movie is a multiplier for the time scale. The rate controls how fast the movie plays. When it is 1.0, the movie plays at its normal speed, meaning that for each second of play, the movie's time advances a number of units equal to the time scale. If the movie's rate is between 0.0 and 1.0 then the movie will play back in slow motion and fewer time units are counted off per second of play. A negative rate implies that the movie is playing backward and a rate of 0 means that the movie has stopped.

The *time value* of a movie indicates where we are in the movie being played back. A time value is expressed in terms of the time scale. When playing forward from the movie's start, the current time value can be calculated as:

$$\text{time value} = \text{time elapsed} * \text{time scale} * \text{rate},$$

where time elapsed is measured in seconds.

Figure 4.1 below illustrates the concepts of time within a QuickTime movie.

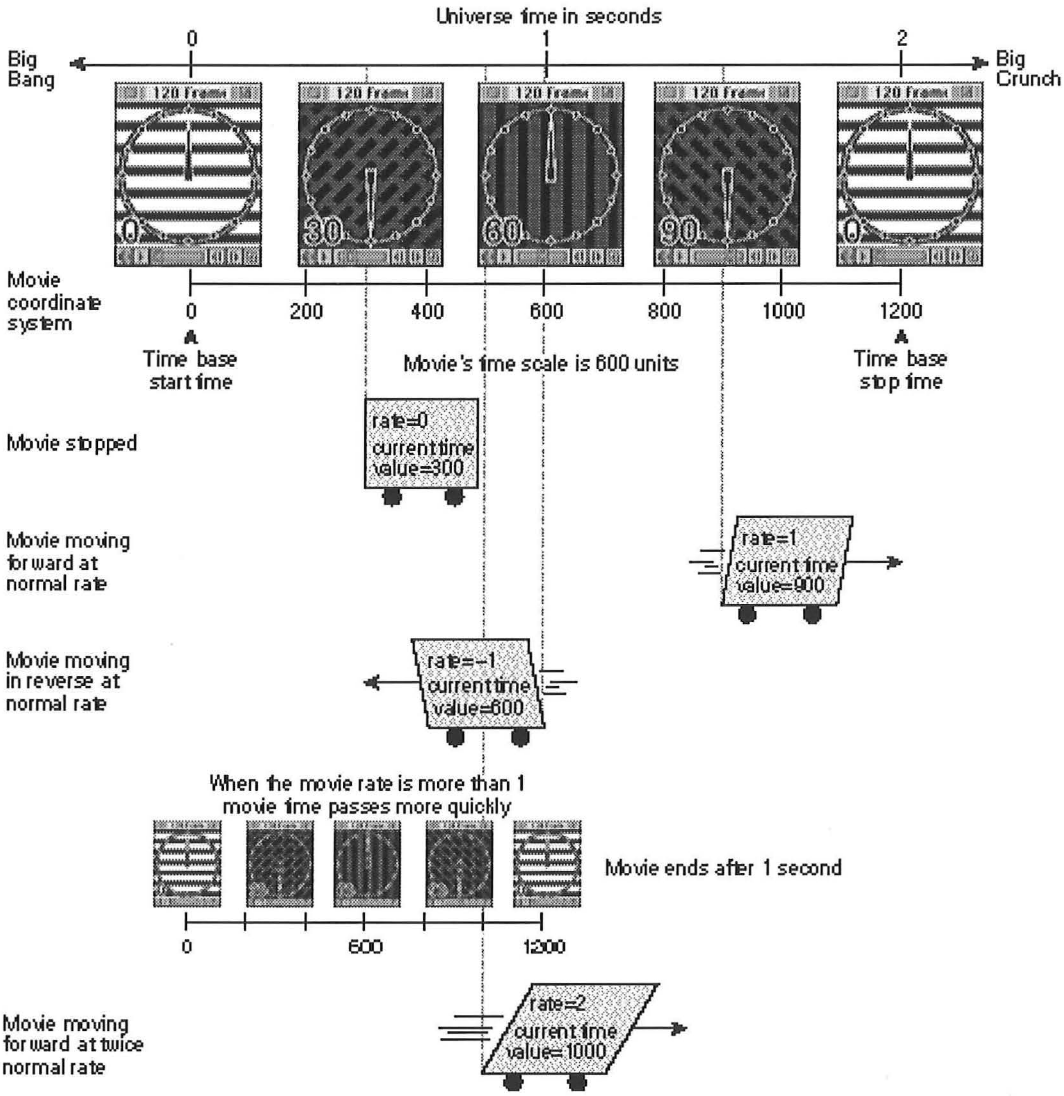


Figure 4.1. Time Concepts in a QuickTime Movie. (From [35]).

The time scale in this example is 600 units. Several examples are shown, showing a movie stopped, at normal speed, in reverse, and one movie at twice its normal rate.

4.2 Architecture of QuickTime

Just as QuickDraw is a standard part of Macintosh system software that allows applications to integrate graphics in a consistent, mainstream fashion, QuickTime is the standard part of Macintosh system software that allows applications to integrate dynamic media in a consistent fashion.

The QuickTime architecture consists of four major components: system software, file formats, compressors, and Human-Computer Interface standards.

- The system level software hides compression algorithms (image compression manager) and external devices (components) from application software, enabling the creation, modification, playback and storage of time based data in a "transparent" fashion. This component of QuickTime ensures access to the best available compression technology and the support of any new hardware component connected to the CPU without modification to existing software programs.

- The movie file format is a container for dynamic data. A movie contains groups of homogeneous data, called tracks. Movies can contain one or many tracks. Initially QuickTime recognises two track types: video and audio. Apple will continue to extend the file format with additional track types such as the text track introduced in QuickTime 1.5. One of the most important features of the file format is its potential for cross-platform development. Any programmer has access to the full specifications of the file format and is encouraged by Apple to develop the appropriate codes to read QuickTime data files from any computer platform.

- The five software-based compression and decompression schemes of QuickTime 1.0 are designed to compress a variety of data from any colour-capable Macintosh. The compressors are highly specific to the type of data they compress. The compressors available in QuickTime 1.0 are: the *Video Compressor*, the *Still Image Compressor*, *8-bit Graphics Compressor*, *Animation Compressor* and the final scheme called *None*, signifying that the scheme does not compress data, but can store and retrieve the information in 1–32 bit formats. Two new video compressors have been added since the initial release of QuickTime 1.0. They are the *Compact Video Compressor* released in QuickTime 1.5 and the *Cinepak* released in QuickTime 1.6.1.

- The Human-Computer Interface guidelines support several elements: the standard movie controller, the standard preview file dialog box, and the other standards for capture, compression and editing.

Figure 4.2 illustrates the main features of each of the four major components of QuickTime.

System Software	1. *Movie Toolbox:* creation, editing, & playback of movies 2. *Image Compression Manager:* shields applications from compression algorithm details 3. *Component Manager:* interfaces applications with external device details (e.g. digitiser cards)
File Formats	1. *Movie:* file format for dynamic data 2. *PICT Extensions:* compression and preview support
Compressors	1. Photo Compressor 2. Animation Compressor 3. Video Compressor 4. Graphics Compressor 5. None 6. Compact Video Compressor 7. Cinepak Compressor
Human Interface	1. Standard Movie Controller 2. Extended Standard File Dialog Box with Preview 3. Guidelines for compression, capture, and more.

Figure 4.2. Components of QuickTime. (From [23]).

The following sections explore each component in greater detail.

4.3 System Software

QuickTime system software consists of three major pieces: Movie Toolbox, Image Compression Manager, and the Component Manager.

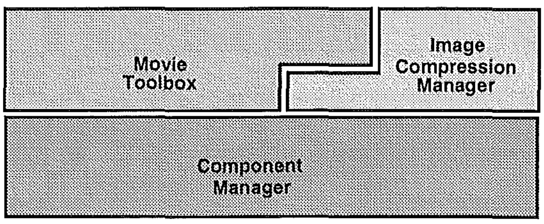


Figure 4.3. Components of QuickTime's System Software. (From [23]).

Figure 4.3 illustrates how the three system software components interact with each other.

4.3.1 Movie Toolbox

The Movie Toolbox is a set of high-level system software services that make it easy for developers to incorporate support for movies into their applications. It provides the methods for storing, retrieving, and manipulating time based data stored in a QuickTime movie.

4.3.2 Image Compression Manager

The Image Compression Manager (ICM) shields applications from the intricacies of compression and decompression through device and algorithm-independent services. The Image Compression Manager allows developers to take advantage of numerous compression schemes (e.g. MPEG, Group 3 fax) in their applications, without having to make modifications to their programs. Section 4.5 discusses *Image Compression* in more detail. The compression schemes can be thought of as plug-in modules that QuickTime can take advantage of if they are available. This ensures that QuickTime-compatible applications will always have access to the best available compression technology. Details of display such as clipping, scaling, crossing screens, and fast dithering are supported automatically via the Image Compression Manager. QuickTime's fast dithering scheme allows the authoring of a movie in 24-bits, and the playback in 8, 16, or 24-bits with the same playback speed.

4.3.3 Component Manager

The Component Manager allows external resources (e.g. digitiser cards, video tape recorders, system software extensions) to register their capabilities with the system at run-time. In QuickTime, Apple defines a class of components for video digitiser cards, compression schemes, and timing clocks, amongst others. Before QuickTime, any application that wanted to support digitiser cards, required custom software for every digitiser they wished to support. With QuickTime, hardware is now transparent to applications. The application merely makes a request to the Component Manager for "a digitiser card with X capabilities" and the Component Manager takes care of locating and communicating with components of that type. In this way, applications can support new hardware without modification.

4.4 Movie File Format

A traditional movie, whether stored on film, laser disk, or video tape, is a continuous stream of data. The term movie in QuickTime refers to *all* dynamic data. For example, a movie can be a presentation slide show, a dynamic bar chart of data, or a graph of lab data over time. The movie file format is a container for this dynamic data. A movie contains groups of homogeneous data, called tracks. A simple movie might contain a video track and a sound track. QuickTime takes care of synchronising these tracks when the movie file is played. The description of the data is separate from the data itself, which allows for multiple edits or versions of the data without duplicating the content each time. Each track refers to its media which contains references to the movie data. The movie data may be stored as images or sound on a hard drive. The data references constitute the track's media. Each track has a single media data structure (figure 4.4).

It is possible to have the movie data stored on multiple devices like CD-ROM disk, floppy disk, and hard disk, while the movie file may be stored on a separate hard disk. Each track in the movie, stored on the hard disk, refers to its media which contains references to the actual data. The media itself may be on any of these other devices and not necessarily with the original movie file.

Unlike a traditional movie that only has one video and one audio track, a QuickTime movie may have any number of tracks. For example a QuickTime movie may consist of five audio tracks and three tracks containing video data.

Each *track* contains time based data such as video sequences or audio sequences. A video track consists of a series of *frames*. These individual frames are still images of video information. Each frame may consist of an entire representation of a picture or may hold the differences between subsequent frames.

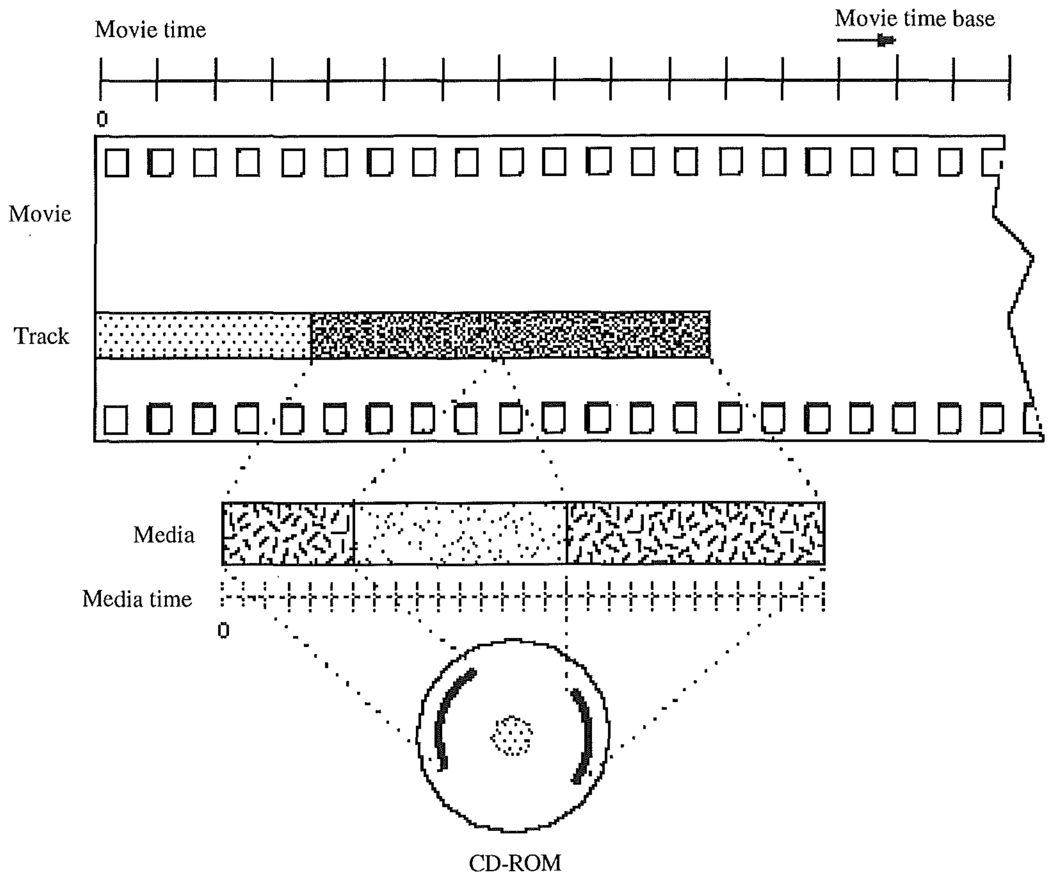


Figure 4.4. QuickTime Movie Structure. (From [22].)

Figure 4.4 illustrates the mapping that occurs of the movie information onto the movie data stored on the storage device. The track in figure 4.4 contains references to its media. The track only refers to a segment at the start and at the end of the media, leaving a segment unused in the middle. The media itself references where it is located on the physical storage devices.

Since the movie itself is not the medium, but the organising principle, valuable disk space is saved. It is possible for many different movies to share tracks amongst each other.

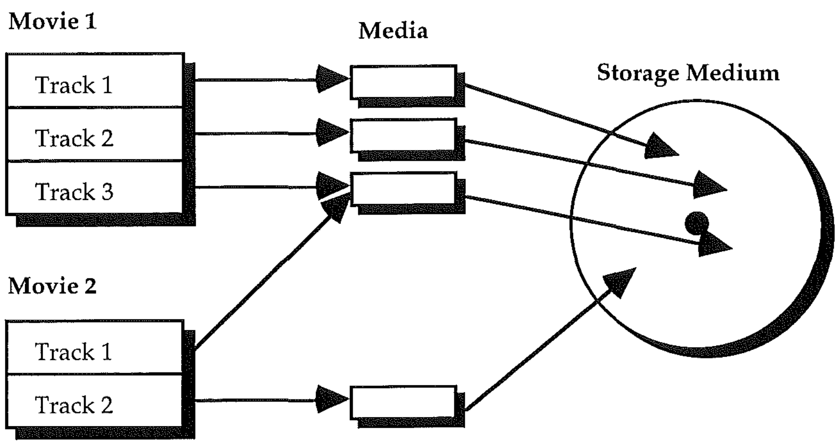


Figure 4.5. Illustration of Movies Sharing the Same Media.

Figure 4.5 shows two movies, Movie 1 and Movie 2, that are sharing the contents of one media. The tracks within each movie are pointers to the media that in turn reference the physical storage device. Since the two movies contain tracks that reference the same media, Track 3 in Movie 1 and Track 1 in Movie 2, the actual contents of the track in each movie are not duplicated. The physical storage space is shared, thus saving storage space.

The movie file format is a cross-platform file format. Apple has published the full specifications of the file format, which encourages third-parties doing cross-platform development to use the file format as a container for exchanging dynamic data.

4.5 Image Compression

The Image Compression Manager provides applications with an interface for compressing and decompressing images and sequences of images that is independent of devices and algorithms.

Uncompressed, full-screen, full-motion video requires a data transfer rate of about 25 MBytes/sec, yet a typical hard drive has a data transfer rate of only 1 MByte/sec and a CD-ROM drive only about 100 KBytes/sec. QuickTime solves this data transfer problem by using software based video compression, requiring trade-offs in the areas of frame size, frame rate, image, and sound quality. Hardware based solutions can improve the trade-off required, but at a cost.

The frame size determines how large the movie will be on the screen. The larger the frame size, the greater the number of pixels that have to be updated for every image displayed.

Figure 4.6 shows some common frame sizes and their respective number of pixels per frame.

Frame Size	Description	Pixels/Frame
640x480	Full-screen	307,200
320x240	1/4	76,800
240x180	1/8	43,200
160x120	1/16	19,200

Figure 4.6. Common Frame Sizes For Digital Video.

To display full screen images, the number of pixels to be updated become an important factor.

The frame rate is another equally important factor to consider. Typically 25 frames per second are displayed in PAL standard television broadcasts. If we reduce the number of frames that we store then we can significantly reduce the amount of information that needs stored and transferred from storage devices. Between 10–15 frames per second is about the lowest threshold for a movie to have an acceptable playback speed for the viewer.

Image and sound quality contribute to the size of information stored. Higher quality images and sound require more space and processor time. The size of the image information can be reduced by changing the number of colours that the image is represented in. For example, instead of each pixel being represented by 16 bits of information, it could be represented by 8 bits. This reduces the number of possible colours for each pixel from approximately 32,000 down to 256. Sound can also be reduced by sampling it at a lower frequency and with fewer bits.

There are many schemes available in trying to reduce the amount of video data to be stored. At any one time video only displays every second scan line on the display. The two sets of scan lines are alternated repeatedly without the viewer noticing. Dropping every second scanline has the effect of halving the amount of data needed to be stored. Another possibility is to reduce the viewing rectangle. The smaller the size of the video display window the less information is needed for each frame. In version 1.0 of QuickTime, the standard screen size was 160 by 120 screen pixels. (1/16th the size of the standard colour screen on a Macintosh).

The second scheme which QuickTime incorporates is compression of the video. Redundant information is discarded as it analyses the colour information in the movie frames. There are two main types of image compression available, *spatial* and *temporal* compression. Spatial compression is concerned with removing redundant information out of static images. Temporal compression removes redundant information out of successive images.

There are many Apple compressors available with QuickTime designed for different purposes. In QuickTime 1.0 there were five compression schemes available; Photo compressor, Animation compressor, Graphics compressor, Video compressor, and None. The Photo compressor is basically a JPEG compression scheme used for still images that can produce compression rates between 5:1 and 100:1. The Animation compressor is designed for computer-generated images with either a lossy or a lossless mode. The Graphics compressor is an Apple developed scheme. It is a lossless scheme for 8-bit still or moving images. The resulting files take up half the space of movies using the Animation compressor. However, there is a trade-off — the playback is at half the speed. The Video compressor is another Apple-developed scheme for storing and playing back compressed digital video. The compression achieved is between 5:1 and 25:1. The last scheme, *None*, provides no compression but still stores the information as a movie in 1–32 bit depth.

Apple has begun demonstrating the integration of MPEG technology into applications using QuickTime at the Tokyo Multimedia Conference (November 22–24, 1993) [3]. The plan is to integrate MPEG into future versions of QuickTime. MPEG offers between 50:1 and 100:1 compression ratios.The quality of video compressed with the MPEG algorithm at rates of about 1.2 Mbits/s has often been compared to VHS recordings. HDTV compressed video is between 20-40 Mbits/s as a comparison.

4.5.1 QuickTime Compression Schemes

In QuickTime, version 1.0, Apple provides a basic set of software compression/decompression schemes that meet a range of compression needs for still images, animation, and video. These schemes are designed to playback with reasonable performance on any colour-capable Macintosh. These compression schemes are a built-in feature of every colour-capable Macintosh using QuickTime. New software or hardware-based compression/decompression schemes can be added to the system by merely adding a compression/decompression (codec) component file to the System Folder. There are five software-based compression and decompression schemes that initially came with QuickTime 1.0 and several more have been introduced since. The basics of each Apple compressor is listed below.

Photo Compressor

JPEG (Joint Photographic Experts Group) [36, 45] is designed to compress full-colour images and typically give compression ratios in the range of 10:1 to 25:1 with no visible picture degradation. Apple's implementation of JPEG complies with the ISO (International Standards Organisation) baseline standard [15], and is a lossy scheme—a compressor in which image data is lost every time the compression is performed.

Animation Compressor

Apple's animation compressor employs a compression algorithm based on run-length encoding principles to compress computer-generated sequences from 1–32 bits in depth. The algorithm supports both lossy and lossless modes. Compression ratios vary widely based on content. The Animation Compressor displays animations at acceptable speeds on low-end Macintoshes, and allows complex animations to be previewed on a Macintosh without first having to lay them off to videotape one frame at a time.

Video Compressor

Apple's Video Compressor employs an image compression method developed by Apple. This compression scheme allows digitised video sequences to decompress from a hard disk or CD-ROM in real-time with no additional hardware on any colour-capable Macintosh. Compression ratios typically range between 5:1 and 25:1. An average movie compressed with this scheme can play back at 15 frames-per-second at 160x120 pixels on a Macintosh IIsi, although the scheme is not restricted to this size.

Graphics Compressor

The Graphics Compressor also employs an image compression method developed by Apple. It provides lossless compression of 8-bit images and is ideal for compressing both still images, such as those created in painting applications, and 8-bit movies. The Graphics Compressor differs from the Animation Compressor, which can also compress 8-bit data, in that the Graphics Compressor gains compression at the expense of decompression speed. A movie compressed with the Graphics Compressor will usually be half the size of an Animation Compressor (RLE) movie, but at approximately half the maximum playback rate.

None

This scheme does not compress data, but can store and retrieve the movie information in 1–32 bit format. It is provided in the situation where no compression is desired.

Compact Video Compressor

The Compact Video Compressor was released with QuickTime 1.5. It provides twice the playback area at the same frame rate, 240x180 pixels compared with 160x120 pixels.

Cinepak Compressor

This codec was released with QuickTime 1.6.1 providing vastly improved compression and playback capabilities.

Motion Pictures Expert Group (MPEG)

This is yet to be released as a compression option for QuickTime. It is planned to integrate MPEG [12, 24] into future versions of QuickTime [46].

4.5.2 Principles of Video Compression

The are two main types of compression in QuickTime: spatial compression on individual frames and temporal compression which is applied to sequences of frames.

Spatial compression reduces the information in each image by removing redundant information. Some of the techniques employed take advantage in the limitations of the human eye, such as having trouble detecting small changes in colour. Colours that vary in very small amounts can be ignored and treated as one block, which makes it easier to compress.

There are two main types of spatial compression, lossless and lossy. Lossless compression preserves all of the original information when it is encoded. The original information can be reconstructed without any details missing. An example of lossless compression performed in another field is text compression. It is no use to trying to compress text if the original information cannot be retrieved. The amount of compression is less than is achievable with lossy compression, such as is used by the JPEG compression scheme.

Lossy compression does not retain all of the information when compressing the original material. There is usually redundant information in image data that can be removed that will not affect the overall effect of the image. The method used by QuickTime separates the image's RGB information into three components: luminance, or light intensity, and two chrominance components, which contain the colour data. JPEG supports lossless compression, through differential pulse-code modulation (DPCM), and lossy compression through discrete cosine transformation (DCT) methods. The DCT method was designed to have a controllable quality based on the quantisation method chosen, and the DPCM method was chosen to give a lossless or near lossless compression [40]. The baseline system, the minimum requirements of the JPEG standard, uses a principle of discarding and compressing using an 8 bit encoder [45]. DCT-based compression works on a stream of 8 x 8 blocks of grayscale image samples. Colour image compression can then be approximately regarded as compression of multiple grayscale images, which are compressed entirely one at a time, or by alternately interleaving 8 x 8 sample blocks from each in turn. Each 8 x 8 block is input into the compression model, yielding output in a compressed form [45]. Lossy compression can achieve significant improvements in results over lossless methods. This however is at the expense of losing some of the original information. If the information needs to be restored to its original state then lossy compression is not appropriate. JPEG was designed originally for still image compression. It has been implemented for movies with three levels of compression quality: standard play (SP), long play (LP), and draft.

Temporal compression, also known as frame differencing, is a technique used by QuickTime, and other digital video schemes, that stores and updates only the pixels that differ from the previous frame, so that much less data has to be stored and displayed. Figure 4.7 illustrates the technique of frame differencing with the image on the right containing only the information that has changed from the previous frame on the left.

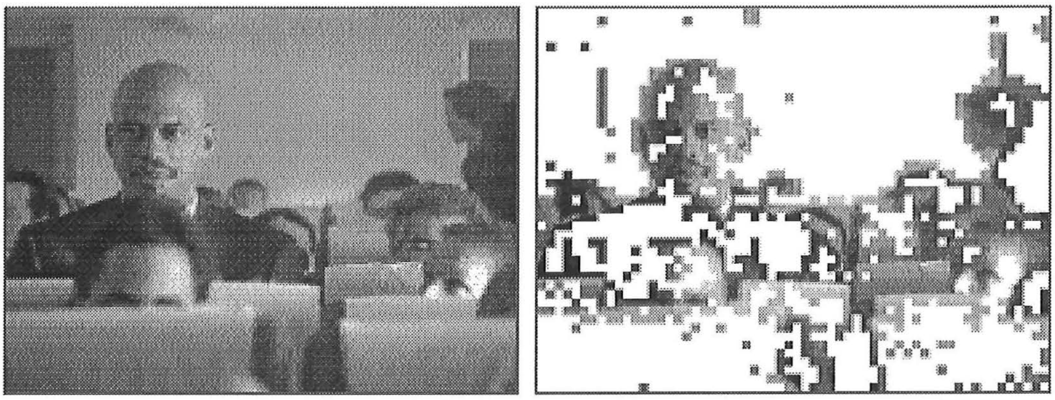


Figure 4.7. Example of Frame Differencing. (From [34]).

As a consequence of this technique, less data needs to be stored on disk for the second frame and it takes less time to update the image. This allows larger frame sizes and rates which produce better quality movies. This differencing of frames can not be done indefinitely. At regular intervals a 'key frame' — a frame that refreshes the entire movie area, and not just the pixels that differ from the previous frame — is inserted. This normally occurs once per second as a rule of thumb. Images with constant backgrounds and very little movement will produce good compression results. Video noise, as one can see when a television set is not tuned properly, increases the amount of information that changes between frames. If the video source is very noisy then the majority of the information has to change between frames, reducing the effectiveness of frame differencing. A clean video source is always recommended to achieve the best compression results.

The audio component of a QuickTime movie is a major contributor to the total size of a QuickTime movie. It is common for the audio tracks to take up to half of the total size. There was no compression of the audio tracks available until version 1.6 of QuickTime. The only means of reducing the amount of information required by the audio track before this was by sampling the audio tracks at a lower frequency. By reducing the sample rate of the recorded sound, the size of the audio information is reduced and the playback of the entire movie may be improved slightly since less information has to be processed. This is at the expense of the quality of the audio.

When capturing the information, the tracks are often physically stored one after another in the movie file. Figure 4.8 shows an example of the tracks being stored one after another in the movie file.

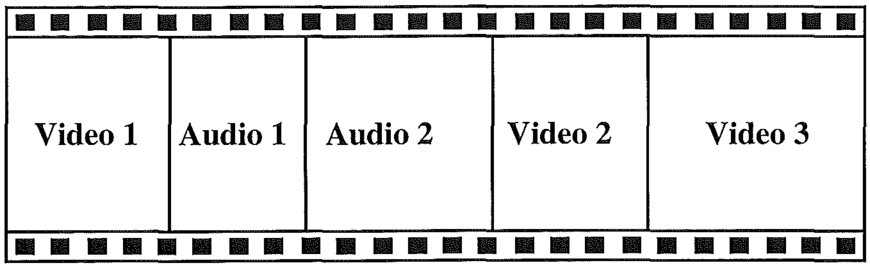


Figure 4.8. Location of Tracks Stored in the Movie File.

To play back the movie a large buffer must be allocated for each track with a lot of seeking between tracks occurring during playback. This leads to poor playback performance. To solve this problem, QuickTime offers the capability to interleave the video and audio tracks (figure 4.9).

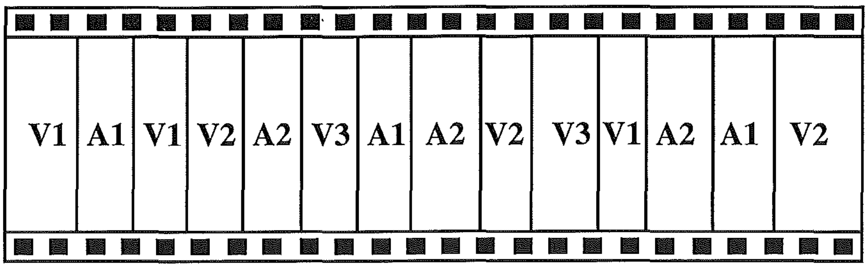


Figure 4.9. Interleaved Tracks in a Movie File.

Since the video and audio tracks are now close to each other in the file, seeking is minimised and only one buffer is required.

4.6 Component Manager

QuickTime provides components so that every application doesn't need to know about all possible types of audio, visual, and storage devices. A component is a code resource that is registered by the Component Manager. The component's code can be available as a system-wide resource or in a resource that is local to a particular application. The Component Manager provides applications with access to external services without the application needing knowledge of the resource providing the service. An application can ask for a device with specific capabilities and expect the Component Manager to locate and communicate with a corresponding device, if one exists.

For example, if the data being played back in a movie represents a compressed image, the Image Compression Manager is called upon to decompress the data into a usable form. The Image Compression Manager, in turn, asks the Component Manager for an appropriate decompression component which may be hardware or software.

The Component Manager controls any number of system-wide resources and those local to an application. When an application needs the services provided by a component the Component Manager allows the application to connect to the component. All components written adhere to an interface so that components can be easily interchanged. An application can chose any component that is available and know that it will work. Some of the components available for QuickTime include the movie controller component, the standard compression dialog components, image compression components, sequence grabber components, video digitiser components, movie data exchange components, derived media handler components, and the clock components.

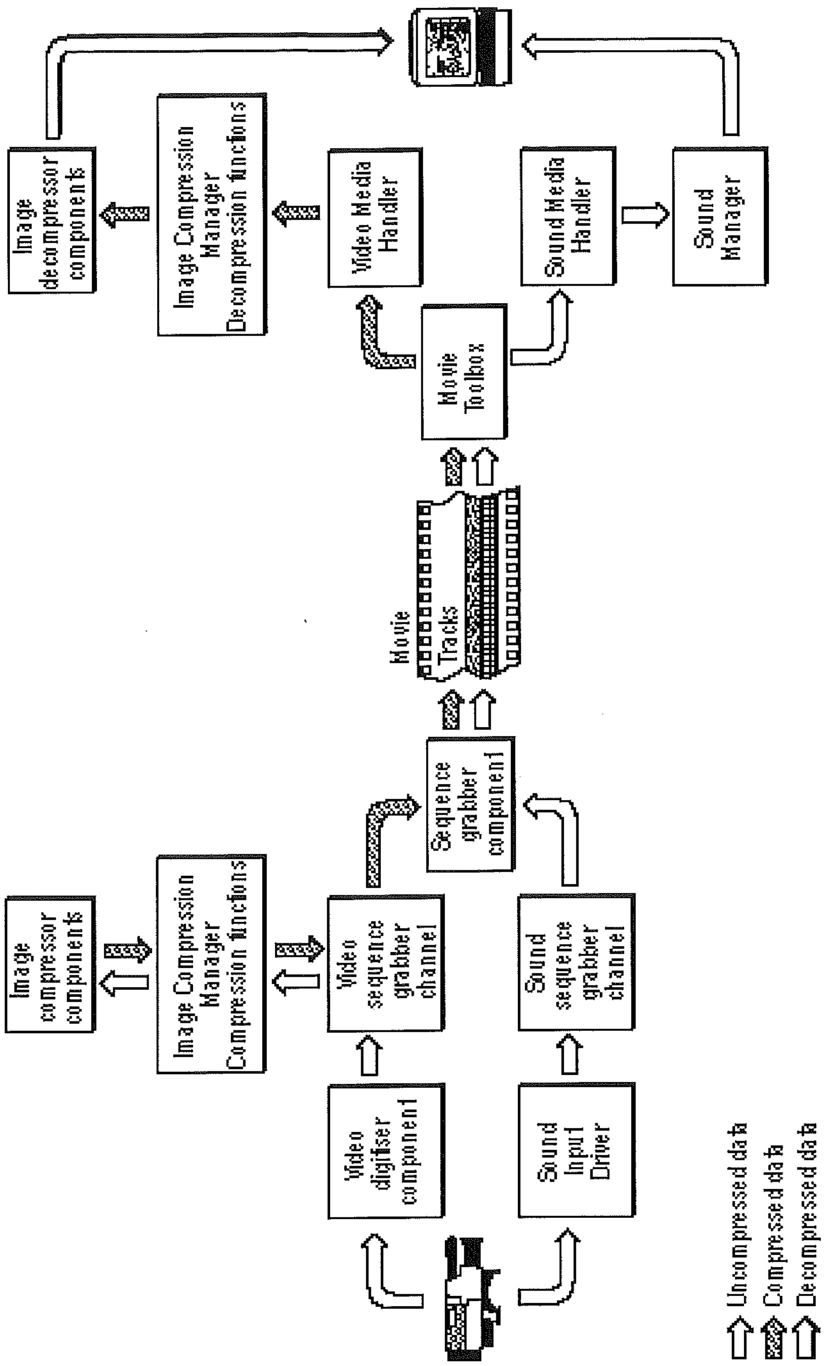


Figure 4.10. Relationships of QuickTime Components. (From [22]).

Figure 4.10 illustrates the relationship between the QuickTime components for recording and playing back QuickTime movies. The left hand side of figure 4.10 represents the components responsible for recording movie files adn the flow of data, while the right hand side represents the components used to play back movie files and the flow of data amoungst them.

4.7 Human Interface

Just as Apple specified human interface guidelines for the original Macintosh to provide consistency between applications for dealing with graphics, Apple has provided the same with QuickTime for dynamic media. In QuickTime 1.0, Apple has specified several standard human interface elements. These human interface elements are the standard movie controller, standard file dialog with preview, and guidelines for compression, capture and other digital video elements.

4.7.1 Standard Movie Controller

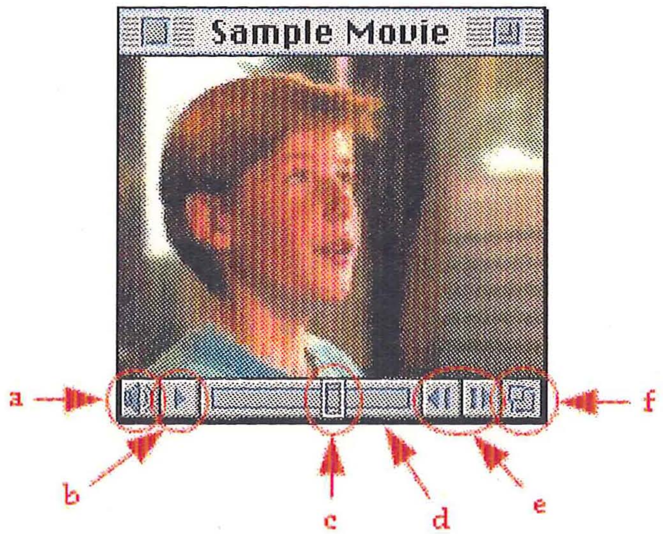


Figure 4.11. Standard Movie Controller.

The movie controller is a system software component that provides a consistent way of controlling movies. Figure 4.11 shows the movie controller that is available to all applications. It provides all of the basic playback facilities needed for controlling a movie. The standard controller allows the user to:

a) adjust sound volume

b) play or stop a movie

c) get an indication of where in the movie you are

d) interactively jump and scroll around the movie by moving the indicator, (c), to any location in this area

e) step-forward and step-reverse through the movie

f) resize the video play back area

4.7.2 Standard File Dialog with Preview

The extended standard file dialog box is another standard system call, provided by QuickTime, that allows applications to invoke a dialog box (e.g. for the Open... menu item) that includes a preview window for still images and movies.

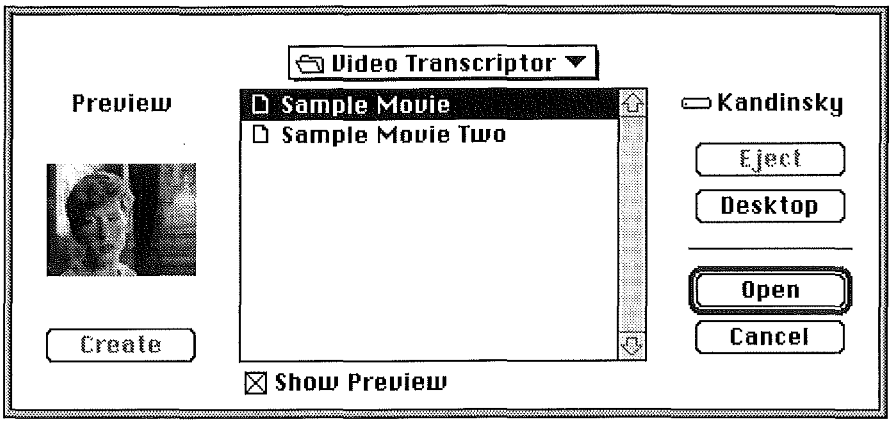


Figure 4.12. Standard File Dialog with Preview.

Figure 4.12 shows the standard file dialog box with the extension to the left hand side of it. This area contains the small preview window containing a still image representation of a movie or a static image.

4.7.3 Guidelines for Human Interface Elements

In addition to the standard elements provided, Apple Computer has released human interface guidelines for areas such as capture, compression, and editing. These elements consist of specific dialog boxes that can be called from any application developed to supply a specific function. For instance, the compression element contains the setting for the particular compression scheme and settings that a user is able to select from. The capture interface allows the user to choose where the source of the material is coming from — a digitising card, CD-ROM, laser disk, or other source.

4.8 A History of QuickTime

QuickTime has had four official releases since its initial announcement in December of 1991. The following is a brief overview of the features that each release has to offer.

4.8.1 QuickTime 1.0

QuickTime 1.0 was officially announced to developers and customers on December 10, 1991 [20]. The beta version of the software had been available for some time before to registered developers. This was seen as the way for handling new and exciting types of information at an easy to use level. Roger Heinen, Apple's vice president and general manager of Macintosh Software Architecture division, said:

> *"We believe that QuickTime will spawn a whole new era in personal computing" , "In 1984, Apple introduced users to the combination of text, and graphics. Today we are extending the combination to include video, sound and animation. QuickTime combined with exciting new third-party innovations will provide Macintosh users with powerful new functionality while maintaining the simplicity and consistency users have grown to expect." [20].*

The features that QuickTime 1.0 offer have been discussed in Section 4.1 — *What is QuickTime*?

4.8.2 QuickTime 1.5

Released on October 19, 1992 [1], QuickTime 1.5 featured better software compression and decompression capabilities, improved playback from CD-ROM, better network support, Photo CD support, text in movies, one-bit display support, and an expanded human interface. These features include:

- larger software-only compression video windows. The new Compact Video Compressor provides the same playback for images twice the size (240x180 for QuickTime 1.5).

- improved compression and decompression, achieved through a new codec (compression/decompression system extension) called the Compact Video Compressor. This allowed movies with twice the frame area to be played back at the same speed as in QuickTime version 1.0. The new size of images supported was 240x180 pixels for QuickTime 1.5 compared with 160x120 pixels for QuickTime 1.0.

- better playback from CD-ROM. Enhancements were made to the low level data handler which optimised data transfer from CD-ROM together with the support for double-speed drives.

- better network performance. The data handler was optimised to perform better over an Ethernet network running AppleShare. With AppleShare 3.0 and QuickTime 1.5, movies will play back faster.

- support for Kodak Photo CD through use of the Photo CD decompressor.

- improved playback on PowerBooks. A special dithering feature has been implemented which provides the full frame rate playback of colour movies on 1-bit screens.

- the expanded human interface that supports the look and feel of system 7 windows.

- support for creating and playing text tracks. Text tracks give the ability to display words for variable lengths of time. For example, subtitles used in foreign movies. The text tracks are interpreted by the Text Media Handler, which comes with QuickTime 1.5. Text tracks behave the same as the current video and sound tracks. Support for generic media handlers that let developers create new movie-track types.

- support for black and white monitors via a special dithering algorithm.

- support for the new Standard Compression Component and the Preview Viewer Component.

- support for the Sequence Grabber Panels that allows a standard way to select compression, source and sample settings for video and audio capture channels.

- new movie input and export components. This allows the conversion of pictures and sound directly into QuickTime movies.

- support for digitisers that include hardware compression.

4.8.3 QuickTime 1.6

In April/May 1992, Apple introduced QuickTime 1.6. This release offered many new features, including:

- the QuickTime 1.6 system extension only taking up about 20K of memory when it is not in use instead of 80K.

- the ability to convert tracks from audio CDs into movies with the assistance of the new AppleCD 300 CD-ROM drive.

- support for the Macintosh Easy Open system extension. This extension is capable of translating sound and image files from a variety of formats into QuickTime format.

- support for ColorSync. The ColorSync control panel is designed to obtain accurate colours from compressed still-image files, for all the devices connected to the Macintosh.

- support for 4-bit grayscale dithering, providing a wider range of shades of grey.

- significant speed improvements.

- audio compression.

4.8.4 QuickTime 1.6.1

QuickTime 1.6.1 was released in June/July, 1992, to fix some minor bugs and to offer better playback performance. The major feature of this release is the Apple Compact Video compression-decompression (codec) component called Cinepak providing vastly improved compression and playback capabilities.

4.9 QuickTime on Other Platforms

QuickTime is not restricted to one platform but has progressed onto numerous other vendor's machines. This progression of QuickTime includes both the conversion of source material from different platforms into QuickTime movies and that ability to play back movies on different platforms.

On January 13, 1991 [19] Apple Computer announced the release of the Movie Exchange Toolkit that allowed developers to convert data on other platforms to QuickTime movies. Sequences of images from a machine can be converted into a time based movie that can be played back on any machine that is capable of playing QuickTime movies. Apple has openly released the file formats of the QuickTime movies encouraging cross-platform development. This open approach by Apple has assisted in making the QuickTime movie format a widely accepted standard on multiple platforms.

Also in January 1991 Apple also demonstrated a prototype of a QuickTime player for Windows at the MacWorld expo. Since this time Apple Computer have been fully supporting the Windows environment with the Windows Software Development Kit that is available for QuickTime through the Apple Programmers and Developers Association (APDA). This provides both sample movie players and the tools necessary to develop Windows QuickTime Savvy products.

Many different hardware vendors now have software available on their platforms to support the playback of QuickTime movies. Many QuickTime applications that have been common on the Macintosh are also being ported to the PC platform, showing the dedication from software companies to the QuickTime standard. For example, Adobe Premiere, one of the most popular pieces of video editing software on the Macintosh, has been ported to the PC market.

QuickTime players are also available on UNIX Workstations. There are several commercial systems available. Xanim, written by Mark Podlipec, is a freely available program designed for animation and video playback. His latest version, 2.64.0, released in February 1994, supports the playback of certain types of compressed QuickTime movies. This opens up the development of QuickTime onto the UNIX platform more freely than before. However the performance of the playback is very slow and at present there is no sound support, although it does support multiple video tracks. Atari computer are also rumoured to be developing a QuickTime player for their range of personal computers.

4.10 The Future of QuickTime

> *"In 1984, Apple Computer, Inc. created a commercially successful phenomenon called the Macintosh computer that paved the way for the integration of graphics into mainstream applications. In 1991 we delivered both System 7 and QuickTime, which together will provide a platform for a whole new*

generation of applications, and extend the capabilities of the mainstream applications we use every day."

- Apple Computer, Inc. [23]

When QuickTime is integrated into the Macintosh operating environment, a whole new level of functionality becomes available. This new functionality has opened new horizons as to how computers can be used in the everyday lives of people. Digital video on the desktop is becoming as standard as the copy and paste metaphor became after the introduction of the first Macintosh computers in 1984. There are many uses for digital video: video conferencing, security systems, watching television on the desktop, and using it to analyse human behaviour.

Random-access, digital video editing is efficient because there is instant access to any segment of a movie or segment, because the digital movie can be copied numerous times with no loss of visual integrity, and because the digital data is accessible from any application on the desktop and can be manipulated in the traditional cut, copy and paste metaphor.

Macintosh developers are already incorporating QuickTime functionality into their next-generation products. Some of what they are doing includes the ability to compress and display high-resolution still images in many applications, video conferencing over networks, and annotating documents with audio and video notes. Other more basic features include the ability to cut and paste movie files to any other application. The standard methodology of playing back a movie with the standard movie controller has made the introduction of digital video even easier. Once the principles have been mastered users are familiar with them in any application that uses the standard interface. This standard interface, however, does not provide functions for everything required and no guidelines are available for developers to follow.

Apple has a strong commitment to the future of QuickTime. The dedication to the development of QuickTime can be seen through the improvements in the releases since version 1.0. Its design is such that anyone can design new features like software-based compression and decompression schemes that can just "plug-in" into the QuickTime system architecture. Even more advanced hardware features can be added so that QuickTime can take advantage of them as they become available giving the user the best technology as it becomes available.

New application categories that are becoming increasingly feasible are video conferencing, low-cost movie editing tools, and dynamic CD-ROM magazines. There is an ever-increasing abundance of CD-ROM titles being produced offering software that utilises QuickTime. CD-ROM offers the best solution for QuickTime's main requirement — storage space. The low cost of producing titles has made the CD-ROM industry boom in the last few years. A large number of CD-ROM's being produced utilise QuickTime. Educational software, games, old movie clips, still image catalogues, interactive multimedia, and many more are currently available.

The true potential of digital video is realised when an end user sits down in front of the Macintosh and chooses dynamic data to communicate ideas. QuickTime movies will be useful in multimedia publications, training materials, electronic presentations, electronic mail, and virtually any Macintosh document or application. The only prerequisite to playback a QuickTime movie is the extension that is required in the system folder and a fast enough computer. To capture and edit original material the end user also needs a video source, either a television or a video camera, a digital video frame grabber, and a QuickTime editing software program.

On February 9, 1994, Apple Computer, Inc. announced that a new version of QuickTime — version 2.0 — would be available in the middle of 1994, [46]. This will provide larger video at faster frame rates, support for music, and support for Interactive TV applications of the 90's [28].

Frame rates of 30 frames per second are quoted at a resolution of 320 x 240, or full-screen video (640 x 480) at 15 frames per second on an LC475 Macintosh. Full support for professional-level video editing is available allowing support for time-code, 60 fields per second video. Throughput of data at greater than 3 MB per second is also quoted.

Support for the MPEG is finally to be included. MPEG, by itself, allows only playback. QuickTime 2.0, provides support allowing the user to edit, search for, interact with, and play back the video information.

The final addition is the support for a new track type called the "music track". This is supposed to provide music capabilities with reduced disk storage required. A 300 MB audio track could be reduced to 800KB through the music track.

Chapter 5

The Design of a Digital Time Based Analysis System

The current research has been undertaken to explore the effects that digital video can have on time based analysis methods. In Chapter 3 the world of digital video was explored with a look at what the major computer companies have to offer. Chapter 4 examined Apple Computer's contribution to the world of digital video, QuickTime.

This chapter looks at the design of *Video Transcriptor*, a digital video time based analysis system. It begins with a look at the requirements of the system followed by a discussion on the design and initial testing of the system. Finally, Human-Computer Interface issues concerning *Video Transcriptor* and digital video are discussed.

5.1 Requirements

Research students within the Psychology Department at the University of Canterbury are trying to understand the emotional state of a person by observing their facial expressions, body movement, and their voice. Controlled experiments are prepared for volunteer subjects to undertake. These experiments involve subjects isolated in a room watching specially prepared material on video tape. The contents of the video tape is intended to invoke special responses from the subjects. The subject's facial expressions and body gestures are recorded onto video tape by one visible and one hidden camera. The video tapes of the subjects are analysed to try to understand the emotional state of the subject. This analysis phase involving video tapes takes a significant amount of time.

For example, the time required to analyse one behaviour on a three minute section of video tape information can be anywhere between forty minutes and one hour. The nature of the behaviour and the frequency of events that occur directly affect the time needed for analysis. This is a significant amount of time for analysing a single behaviour. Normally there are several behaviours that are analysed for each subject. This time consuming process has dissuaded some researchers from using video tape as a part of their research.

The majority of time taken is not writing down notes when analysing the video tape information as one might expect, but in the response the video tape player takes. The researcher typically plays a few seconds of the tape, stops it, and then rewinds it to view the section of video tape again. This process is repeated until the information is accurately transcribed. Finding the exact location of behaviours is very difficult when using video tape. It is very common to miss the exact time of the behaviour to be recorded requiring the researcher to rewind the tape to try again.

If the researcher analyses a five second section of tape repeatedly, the following times are typically incurred:

- 2 seconds for the tape to start playing
- 6 seconds for viewing the tape at normal speed (1 second before the actual time)
- 2 seconds for stopping the tape
- 5 seconds to rewind the tape and stop

These are estimates of the times that it would take for playing a short section of video tape. The times were taken using a NZ$2,500 Mitsubishi E-70 video tape player which is faster than the average video tape player. From the figures we see that it takes twice as long to reposition the tape each time it is viewed than the time spent viewing the video tape contents. This process normally occurs several times and then the researcher may use the frame advance to position the video to the exact location in order to record the time accurately. This could typically take ten times longer than playing the video back at normal speed during the section.

If we assume that the researcher only plays the video tape section through twice and then single steps through the section once, we can calculate the time for analysing five seconds of video tape to be

2 x (15 seconds playing at normal speed) +
1 x (stepping through the video tape section – 69 seconds)
= 99 seconds to analyse one behaviour in 5 seconds of video tape.

The final thing is to add the time actually spent recording the event. The time that the behaviour occurs is displayed on the video tape itself. It had been recorded onto the video tape when the subject was originally videoed. When the researcher records the time on the video tape they have to work out the offset from the start of the section that they are analysing. For example the original video may contain thirty minutes of a subject but the researcher is only analysing three minutes of it. The timing information may read eight minutes at the start of the section being analysed. The researcher typically has to calculate the offset for each behaviour that is recorded.

If we estimate that the time spent recording the behaviour is approximately eleven seconds, we can calculate the total time spent analysing a section of video tape that contains a behaviour. This comes to 99 + 11 = 110 seconds, or twenty two times the actual time of the video tape section being analysed. This is only an approximation of the time under the right conditions. The time required may in fact double if it is a difficult section being analysed or may be shorter of it is easier to record or it doesn't require any further analysis. The minimum time that one could expect to record an event would be

5 seconds (for viewing the total section) +
2 seconds (for stopping the tape at the exact location in the section) +
2 seconds (for the tape to start playing again) +
11 seconds (to record the event)
= 20 seconds.

When compared to 110 seconds for an average analysis time of a five second section of video, this is 5.5 times faster for the best case scenario, yet it is still four times the length of the original video section. Based on this, it would take sixty six minutes to analyse three minutes of video tape. However this would imply that the entire three minutes would be analysed in detail. In practice only portions of the video are analysed. Typically researchers were taking between forty minutes and one hour to analyse the section of video information.

A common task performed that video tape technology is unsuitable for is the comparison of different subjects. Commonly the video segments required are located on different sections of the tape, or on entirely different tapes, causing significant delays in trying to compare one subject's video against another. There is no easy way to make this easier for the psychology research students short of simultaneously playing two video tapes side-by-side on different video tape players. The researcher has to slowly go through the video tape taking notes about events that occur and often trying to make comparisons against other subjects. This can introduce significant inaccuracy into the results. For instance, by the time they get to watch the remaining sections of tape their frame of mind may easily have changed during the long delays making the researcher's perception of the picture unreliable. With more money, multiple video tape players could be synchronised to view the subjects simultaneously. This is a very expensive option that is not feasible under the current research. Because of this, the vast majority of comparisons are made against the secondary source — the notes that have been taken. It would be more accurate if all the comparisons could be made against the primary video tape information.

Psychology research students accumulate a vast quantity of information rapidly that is very difficult to process on paper. This information is conventionally entered into statistical analysis packages for further processing to find results. This is a two stage process; firstly writing the notes down on paper, which occurs when analysing the video tape information, and secondly entering them onto the computer.

Analysing the material on video tape can be a time-consuming and relatively difficult process. Section 1.2 in Chapter 1 discusses the experiments and the analysis methods currently being undertaken in the Psychology Department.

The Psychology Department approached the Department of Computer Science for assistance with this problem. The basis of the research that has been carried out was to see how computers could be utilised to assist in the analysis of video tape information. Research into previous and existing analysis methods showed that the majority of computer-assisted systems still used video tape players in the process. Chapter 2 discusses the existing systems that were analysed and the drawbacks that they currently have.

Digital video has recently become common-place in the personal computer market. It introduces an alternative method to video tape for storing and playing back video information. One of the major benefits digital video has to offer psychology research students is the instant access to any location within the video information unlike the delays with video tape technology.

The features that need to be addressed for a digital video based analysis system include; easy control of the video information, integrated note-taking facilities, and the ability to analyse the results.

Current note-taking methods involve repeatedly stepping through the video tape, taking notes down as they occur onto paper so they can be processed at a later date. This makes it difficult to analyse the results since there is no longer any relationship with the original video. The computer can offer many advantages for the note-taking just as it has with previous systems, [6, 8, 25, 31, 39]. The note-taking process should be able to cope with many fields of research, being as flexible as possible instead of being tailored to a specific research field. The notes taken should have a direct relationship to the source of the information so that both can be viewed simultaneously.

More emphasis is required on the actual analysis capabilities than actually having to write down explicit details on each event. Conventionally the time is accurately recorded onto paper for each event that is found. Extra inconvenience is encountered when the time on the video tape has to be converted into an offset from the start of the section being analysed. The note-taking process needs to be just that, note-taking, without having to record times for every event. Another important feature of the note-taking process is keeping notes about behaviours separate from each other and not lost within other behaviour categories.

The research being carried out does not directly concentrate on creating any form of complex analysis tools to process the notes after they have been made. There are many commercial statistical packages that offer sophisticated capabilities. All that is required is a simple means to export the information into a format that statistical analysis packages can read in.

5.2 Design and Initial Testing

The design and testing of a digital time based analysis system has been an interactive task involving ideas and feedback from potential users with their reactions during the development of a prototype system.

The design and testing stages were carried out in conjunction with a research student from the Psychology Department at the University of Canterbury. Initial discussions were held to determine what the current analysis method consisted of. This has been covered in Chapter 1 in the section *Psychology Experiments*. Some of the requirements have been highlighted in Section 5.1 — *Requirements*. From the series of initial discussions the current analysis process was determined. Additional desirable facilities that could take advantage of computer-based technology were also discussed. From this information the documentation for a prototype system was written. This was essentially the beginning of a user manual detailing how a digital video analysis system would function. The twenty page document consisted of screen shots of the proposed system together with fully supporting documentation on how the user interface works together with other functions of the system.

The whole process of designing the system has been an ongoing interactive process. Feedback from the initial design documentation helped resolve misunderstandings and clear up problem areas of the design. After several modifications to the proposal it was finally agreed upon for a prototype of the system to be designed and built. The process of developing a prototype of the system became a very time consuming exercise as will be discussed in Chapter 8 — *Implementation Details*.

Once the prototype had been developed, aspects of the system could be tested to determine interface design issues. Initially the interaction of users with the prototype system was video taped so that it could be analysed later to see any problems with the interface during the session.

Several interface design problems were uncovered during the session. The main problems arose from confusion with some of the interface controls. These highlighted areas in the prototype where the interface could be improved. The analysis of the interface has been an ongoing process right from the prototype stage.

Subsequent sessions of testing the prototype system involved observational and verbal feedback from the user instead of using a video camera all of the time. The main reason for dropping the camera was that the users were able to express issues regarding the interface more freely by this time.

Interface design is one of the most important areas in the development of any device that is to be used by people. This area needs special consideration in computer software to make it as easy to use as possible. A significant amount of time should be spent on the design of the interface so that it is easy for the user to operate rather than for the user being required to spend a lot of time to learn the new system. The facilities of the user interface should also be as functional as possible.

5.3 Human-Computer Interface Considerations

> *"Architecture and interface design have an important goal in common: to create liveable, workable, attractive environments. The principle of 'form follows function' maintains that the form of objects should follow their functional requirements. The architectural movement towards functionalism was initially contested in Victorian England, as the industrial age was trying to hide the impact of industrialisation on people's everyday lives. Le Corbusier was initially*

> *viewed with scepticism when he proposed that a house was a 'machine for living', as opposed to the traditional idea that a house was a shelter. He likened a house to a ship or bridge in its function. Likewise, we are only just beginning to conceive of computers as extensions of our functional everyday lives."*
>
> - Brenda Laurel [27]

Louis Henry Sullivan's (1856-1924) famous words "form follows function", which was later adopted by the functionalist movement, can be seen as a basis of the Apple Macintosh desktop metaphor. The Macintosh draws upon metaphors from the real-world and maps them onto its desktop. For example, in an office situation, before computers were common-place, personal files about clients may have consisted of several sheets of paper placed neatly into manila folders, and housed in filing cabinets. When the contents were required, the staff would go to the filing cabinet, find the manila folder, and typically take it to their desk. From there they would open the manila folder, and examine the documents contained within, returning it to the filing cabinet when they were finished. When the client information was no longer required, the documents, or the actual manila folder itself was thrown out into the rubbish, never to be seen again. Here the visual representation of the real-world situation can easily be mapped onto the Macintosh.

In the Macintosh environment the shape, or form, of the objects is dictated by its own metaphor. Documents, or files, are typically represented by an icon that looks like a piece of paper (figure 5.1). Manila folders and the filing cabinet from the office scenario can be seen in the form of folders on the Macintosh. The office worker's desktop can be seen as the screen area on the Macintosh. The user can place documents and folders on it to work with. Finally, the rubbish bin that is a common feature in the office also exists on the Macintosh for the disposal of unwanted documents and folders. The open window in figure 5.1 represents an open folder containing two documents and two unopened folders.

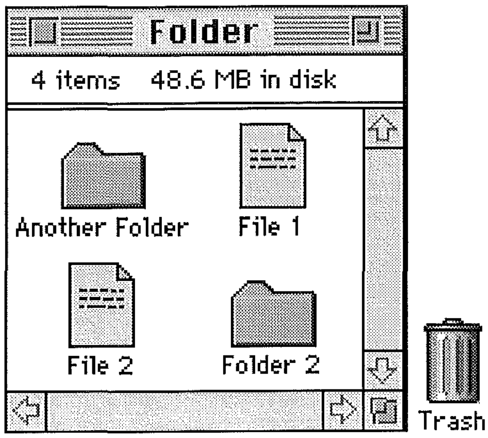


Figure 5.1. Relationship Between the Macintosh Desktop and the Real-World.

Drawing relationships between a computer interface and the real world can assist the user in determining the functionality of the interface elements even easier. The metaphor used with files contained in folders can be directly related to the environment often found in an office.

Hugh honour et al. [18] saw that "Le Corbusier's famous definition of a house as 'a machine for living in' has often been misunderstood and misused in order to denigrate him. He did not mean that a house ought to be an artless, soulless mechanical capsule fit only for robots to inhabit. For him, as for the Russian Constructivists and the Bauhaus architects, the machine was something wholly beneficial. He thought that a house ought to be conceived, designed and produced in a rational manner (as were the motor cars and aeroplanes) and that traditional, irrationally planned and designed houses simply frustrated the promise of the new age and the good life for all that the machine could bring."

Here we can see what Swiss painter-architect Charles-Edouard Jeanneret (1887–1965), known by his pseudonym as Le Corbusier, really meant. Here his notion of the house being conceived and produced in a rational manner can be seen as being a well structured and thought out interface. By contrast, irrationally planned and designed houses are like poorly designed programs, which will not be employed, and the "new age" will not come for these unfortunate users.

Apple realised the importance that graphical user interfaces could have on the computer industry. They invested the time and resources into this field in the early 1980's. This consisted of several members from the learning research group at Xerox Parc, where the famous *Dynabook* project originated [26]. This effort consisted of designing and implementing a new computer architecture strongly based on ideas taken from the Smalltalk environment. This project led directly to the *Lisa* and *Macintosh* microcomputers. The resulting Apple Macintosh and its interface was a revolutionary step when it was first introduced in 1984. The typical interface around this time, and still in common use today, was that of the command-line interface (CLI).

What makes the Apple Macintosh so successful is the overall consistency of its desktop interface, edit metaphor, text manipulation, menu selection, and many other features. This expands to all the broad range of applications written for it by developers who follow a set of guidelines set down by members of the Human-Interface Group at Apple Computer.

Even though the Macintosh user interface may appear to be consistent, there are several flaws to its design. One such flaw is the *Open* menu item found in the *File* menu. When in the Macintosh Finder it will open a selected folder or start a selected application. When the *Open* menu item is used within an open application the functionality changes, it opens documents associated with the application. Another inconsistency in the interface design is when a user empties the trash. The dialog box used to show the trash being emptied used to have a *Cancel* button. This has been subsequently changed to a *Stop* button. The reason for this is that when a user selects the cancel button, the state of the system should be reset to the same state that it was in before the operation was selected. In the case of the trash being emptied any files that were processed between the start of the operation and the time that the *Cancel* button was pressed remain deleted instead of being restored to the original state they were in before the trash was to be emptied [42]. Other inconsistencies exist with the trash can. For example, to remove an application or a document from the Macintosh, the file can be dragged to the trash can where they remain until the user selects *Empty Trash* from the *Special* menu. Dragging removable media like a floppy disk to the trash unmounts and ejects it. Dragging a file-server volume to the trash unmounts the volume. All of these are examples of some of the inconsistencies found in the Macintosh interface.

The central principle of the Apple Macintosh interface is the "noun-verb" action which the user constantly performs. The user first selects an object (noun) either on the desktop or in a window, then chooses, from a menu, the operation (verb) to be applied to the object. Since the menus display the full range of potential activities available, users don't have to remember and type command names. Instead, they can chose from the alternatives presented. Thus the user's task is recognition of the desired action required, not recall of the command to be entered.

Human interaction with computers has changed dramatically since the pioneering days of the first computers. Interaction has changed from the early punched card systems, where response times could be several days, to data visualisation whereby changing parameters will give immediate visual feedback to the user.

Digital video is one area that has come a long way in a short span of time. It was not too long ago when still images were being heralded as the latest advance in the computer world, and now full-motion video is becoming as common-place as still images were several years ago.

Digital video introduces several problems to the area of the Human-Computer Interface, especially in the Macintosh interface. The main problem is the lack of consistent interface guidelines for controlling video. Every vendor is adopting new interface styles, making it more confusing for users who are required to learn a new metaphor for each different application, to achieve the same results. Apple has provided a standard controller interface, as part of QuickTime, that provides basic playback features (figure 5.2).

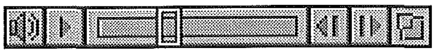


Figure 5.2. Standard Digital Video Controller Component.

It supplies, from left to right, sound volume, play/pause, positioning control to anywhere in the movie, step backwards a single frame, step forwards a single frame, and resizing window control. This standard controller interface supplied is very restrictive in nature for developers. It essentially provides a plug-in controller component for easily controlling digital video information. There is no resemblance to features that users are familiar with, such as those commonly found in the video tape remote metaphor. As a result, many features with this control are not obvious to the user. For instance, to return to the start or end of the video segment, the user is required to hold down the option key and press one of the step buttons. Another hidden feature lets the user change the playback speed of the movie by holding down the control key and clicking in the stepping button area to make a small slider control appears. Developers want to be able to supply the same functionality provided, but in a more obvious manner than some of these cryptic solutions. As a result, many different solutions to controlling digital video have appeared, to provide the facilities that are required.

Brenda Laurel [27], said "The experience of designing real-world objects for everyday use has a lot to teach us about how to design usable interfaces for software. Those designs which endure the test of time are mostly those that are reliable, offering straightforward ease of use and pleasurable user experiences". This suggests that ideas from real-life that deal with similar dynamic data could be useful for reference. Video tape players are a close metaphor to the fundamental underlying structure of QuickTime, and digital video in general, and have been extremely successful in western society. Thus video tape players are the logical place to look for successful interface designs. Since the buttons for controlling video tape players are well known and almost standardised with respect to their functionality, this is an ideal basis for which to take ideas for controlling dynamic data without users having to learn new concepts.

The main set of playback controls settled upon for the current research originate from the remote control metaphor found on video tape players. The video tape player's remote control provides standard features like play, stop, rewind, fast forward and single step facilities. The advantage with this style of control is that the user is able to bring the skills required to operate the main playback controls from their own personal experience. This is the ideal type of interface where the user is not required to learn new material but is able to source their knowledge from elsewhere. This is contrary to the standard movie controller interface set out by Apple. *Video Transcriptor*'s interface is based around existing concepts familiar to users already.

The final design of the prototype system, *Video Transcriptor*, can be found in Chapter 6. More detail of the whole system can be found in the user manual found in Appendix A — *Video Transcriptor User Manual*. This manual is designed to teach new users how to use the system via a tutorial approach. It can be considered just another extension to the interface of the system. Instead of being the part of the interactive system that is used all of the time it is designed to be an additional reference to the system that introduces the user to the system quickly through basic steps. Introducing the user to the new system is done through the *Quick Start* section at the front of the manual, which steps the user through the process of loading a movie, the playback controls, the note-taking process, and finally exporting the results. Users do not like using manuals — in fact it is often the last thing that is picked up. The *Quick Start* section consists of only about six pages with illustrations making it very easy to learn the basic function of the entire system. Once the user has mastered this they can then proceed to learn the system in more detail in three stages. Each stage is one of the three skill levels available, each offering more advanced capabilities as the user becomes more familiar with the previous level.

Chapter 6

Video Transcriptor

Video Transcriptor is the result of the prototype system designed and developed with the assistance of a research student in the Psychology Department. It is a software based note-taking tool for use in the analysis of time based information. It is oriented towards the analysis of video and audio information where video tape technology was used previously.

Video Transcriptor has not been designed for a specific user-group in mind but at a general audience of people who wish to analyse any type of time based information. It is a totally digital based system designed to take advantage of existing technology commonly available. The advantage with this system is the total integration of the time based information with the note-taking facilities on the same computer, instead of using a video tape player to hold the time based data and paper or a computer to store all the notes, [8, 25, 39].

The purpose behind developing *Video Transcriptor* is to see how observers interact with a totally digital based system in their analysis of time based information. From this we should be able to determine if a digital based system offers advantages over traditional methods. The Human-Computer Interface is a major aspect of the design of *Video Transcriptor* because there is no concrete standards for digital video interface design. Many features of the interface design have stemmed from interaction with users during the testing stages.

This chapter discusses the functionality of *Video Transcriptor,* a time based analysis system. The major components are discussed with emphasis on interface design issues and justification of the choices. More details about the construction of *Video*

Transcriptor can be found in Chapter 8 — *Implementation Issues*. A complete description of the system can be found later in Appendix A — *Video Transcriptor User Manual*. The user manual has been a major part in the design of *Video Transcriptor* throughout this research.

6.1 Components of *Video Transcriptor*

Video Transcriptor is designed around the idea of a central controller, the Master Controller, that looks and functions like a remote control unit from a video tape player. This analogy between the remote control and the Master Controller helps the user to feel more comfortable with the controller when they first use it. The Master Controller manages the playback of the video and audio information and provides controls for each of the tracks and windows in a movie. Each video and audio track in a QuickTime movie has a separate window. Each track may also have a separate transcription window where the transcription notes are taken. The three main components of *Video Transcriptor*; the Master Controller, video and audio windows, and the transcription windows, will be examined, emphasising the influences on their interface designs.

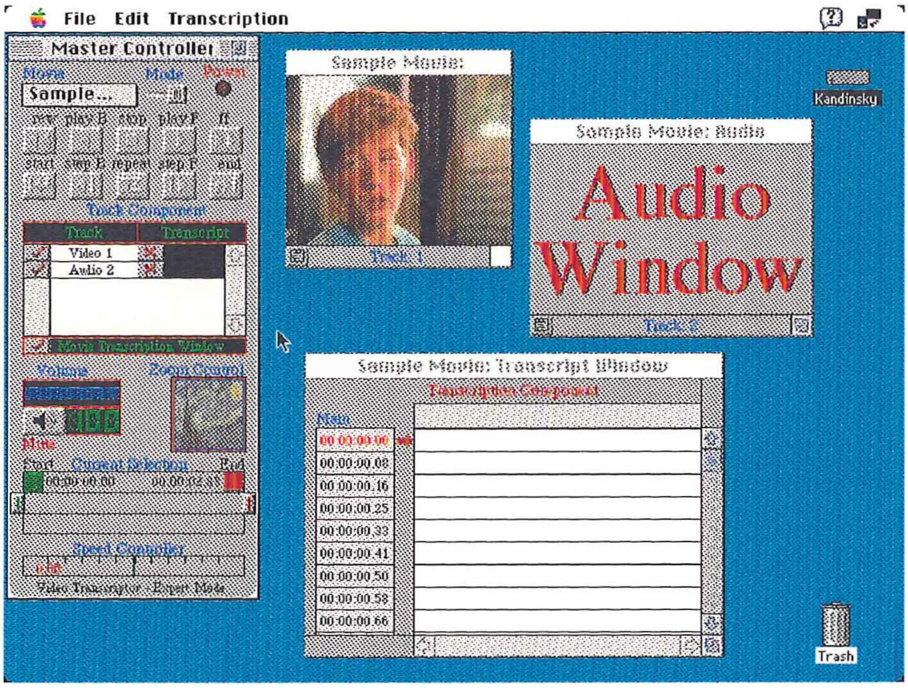


Figure 6.1. Screen Layout of *Video Transcriptor*.

Figure 6.1 shows the different components found in *Video Transcriptor*. They are, clock-wise from top left, the Master Controller window, video window, audio window, and the transcription window. An overview of these different components is discussed next.

6.2 Master Controller

The Master Controller is the central component to the operation of *Video Transcriptor*. It offers a variety of controls for analysing the video and audio information in detail.

The initial design of the Master Controller consisted of a horizontal window taking up the lower third of the screen. This contained many extra features than are no longer present. One such feature removed was statistical information about the number of frames per second that are in a movie. These extra features often distracted and confused the user about the function that the Master Controller was meant for.

The Master Controller changed in design into a vertical window located at the left hand edge of the screen (figure 6.1). The interface was redesigned to look similar to a video tape player's remote control. New users were able to immediately recognise the function of some of the interface components on the Master Controller. Specifically these are the two rows of buttons located at the top of the Master Controller.

User skill levels were also introduced into this new design to alleviate the problem of overwhelming new users with extra features. The new user is presented with a cut down version of the Master Controller, which is the *basic* level. This helps to make the user feel more comfortable when trying to learn the system. As the user becomes more experienced with the interface they may wish to learn more features. This is done by choosing one of the two more advanced levels available, the *standard* or *expert* level, which provide features such as variable playback controls, positioning controls, and zoom control.

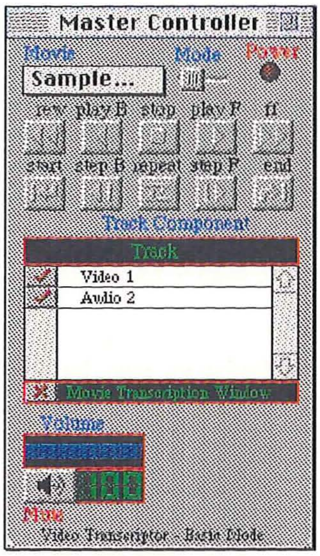


Figure 6.2. Master Controller Window, BASIC Mode.

Figure 6.2 shows the available controls found in the *basic* level of *Video Transcriptor*. The *expert* level can be seen in figure 6.1, which contains all of the controls available. More information about each mode and their differences is available in Appendix A — *Video Transcriptor User Manual*.

The very top of the Master Controller contains a pop-up list of all movies loaded. Researchers commonly need to compare subjects directly or indirectly. By opening up multiple movie files simultaneously it is easy to indirectly compare subjects without affecting what the researcher is currently doing. For example, if the researcher is in the middle of analysing a section of video, they may remember seeing something that may be of interest in another movie file previously analysed. The researcher can load this new movie file to look at, without affecting the analysis already started, or possibly analyse the second video separately.

To the right of this pop-up list is a toggle switch for changing the current skill level of *Video Transcriptor*. This is useful when many different users are using the system for analysis work, each at different skill levels.

The top of the controller contains a set of ten buttons that have a close resemblance to those found on a video tape player's remote control. The functions provided include play forward and backward, stop, step forward and backward, fast forward and rewind, skip to the start and the end of the current selection, and repeat. Most of these buttons provide the basic features that can be found on a video tape player's remote control that the psychology research student is used to. This set of buttons are the focal point of recognition that new users initially have with the system. They are able to relate the functions of the buttons to their own personal experience. The speed of response for these controls is very important. Analysis methods involving video tape players are heavily restricted by their response. In chapter 5 some typical response times for analysing a five second section of video tape were calculated. It was found to approximately take about 110 seconds to accurately analyse a five second section of video tape. Most of this time is spent waiting for the tape mechanism to start or stop. The buttons provided on *Video Transcriptor* for controlling the same functions do not have the same problem — they provide instant response to a users request. This is a significant time saving for the researcher. Other features not provided by video tape methods include the buttons to rewind the movie back to the beginning, or to the end, instantaneously. This is only possible since the digital video is stored on a random access device while video tape is stored on a linear medium.

Below the set of buttons is a list of all the tracks that constitute a movie. This list contains information about whether each track is active and the status of any transcription notes associated with each individual track. There are many times when researchers would like to compare different subjects simultaneously. Current video tape technology makes this difficult, usually relying on very expensive equipment to synchronise special video tape players. *Video Transcriptor* provides the facility for the researcher to do just this, analysing any number of video and audio tracks simultaneously. The list contains an entry for each track in the movie that the researcher can control. Beside each track is a button that activates or deactivates the track. The researcher typically wants to control what they want to see and hear.

At the bottom of the list of tracks is a button labelled *Movie Transcription Window*. When this is activated a window appears for the researcher to begin their note-taking. In figure 6.1, the *expert* level, a second column of buttons is visible. These activate similar transcription windows for individual tracks. Initially new users only have access to one transcription window available. Individual transcription windows for each track in the movie are available when the researcher changes to the more advanced levels.

Below the list of tracks is a volume control and a mute button for controlling the sound of each movie.

Extra controls are available under the more advanced levels of *Video Transcriptor*. Figure 6.1 shows the Master Controller with all the components visible. To the right of the volume control is the Zoom Controller. This is another feature that was considered useful by research students. This gives the researcher the ability to zoom into an area of the video information to analyse a particular feature in more detail. Conventional video tape players have no way to examine a particular area of the screen short of actually viewing it on a larger television set. This analysis facility has been incorporated as the Zoom Controller which allows the researcher to zoom into a specific area of the video windows in the movie.

Immediately below the volume and zoom controllers is an additional control for positioning the movie. This allows the user instantaneous access to any location within the movie. This is one of the major limitations with linear video tape technology whereby several minutes can be wasted while trying to find specific locations of video information. Instant access allows significant improvements in the time spent in the analysis process. As we saw in Chapter 5, the amount of time spent repositioning the video information was far in excess of the time actually spent analysing it.

The secondary feature this control offers is the facility to set a start and end time of the movie for analysis purposes. This is a similar feature that can be found on many audio CD players, usually called A–B repeat. By setting a start and end time the researcher can restrict the area of video and audio information that is currently being analysed. The movie can only be analysed within this period until it is changed. For instance, a two second period may be selected in the middle of the movie and the repeat button selected. By pressing the play button the movie will endlessly loop around this two second section of the movie making it easy to view one section of the movie. Under video tape methods the researcher normally plays the video, presses stop, rewinds the tape, and then plays it again and again.

The last component at the bottom of the Master Controller is for controlling the playback speed of the movie. Video tape normally offers single step and play as the two most common speeds of playback. Expensive broadcasting video editing decks offer variable speed dial controls that would be excellent for controlling the playback speed. More recently the new Mitsubishi video recorders are providing a similar feature on their remote controls. In response to the need for this feature, *Video Transcriptor* provides full control of the movie in slow motion or up to five times its speed in either direction from a slider style control.

6.3 Video and Audio Windows

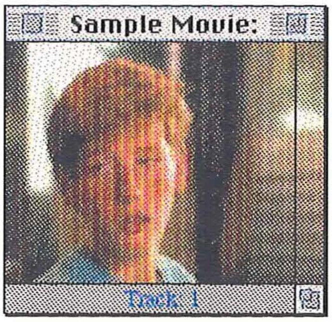


Figure 6.3. The Video Window.

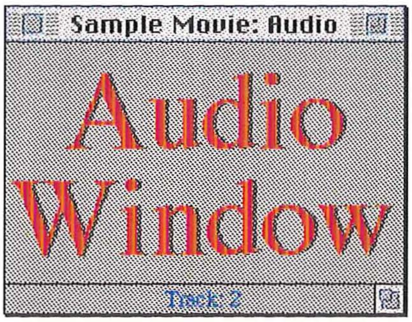


Figure 6.4. The Audio Window.

Figures 6.3 and 6.4 illustrate the video and audio windows that can be controlled with the Master Controller. *Video Transcriptor* provides an individual window for each track in the movie. The implementation of this has seen some significant changes throughout the development life-cycle of *Video Transcriptor*. This is discussed in Chapter 8 in detail. The video windows are resizable but have one side-effect as a consequence; when an individual video window is resized all video windows are changed to the same size as the one that was modified. This is one compromise that had to be made in order to save valuable system memory when using *Video Transcriptor*. QuickTime supports the structure for storing any number of tracks within a movie but does not supply the features to display video information into different windows. Until one of the latest releases of QuickTime, version 1.6, the only way to support multiple video tracks being displayed in individual windows was to have multiple copies of the same movie loaded into memory, all synchronised together. QuickTime 1.6 supports new features that allow only one movie to be loaded into memory, each video track displaying into separate windows. Section 8.5 — *Multiple Track Support*, in Chapter 8 discusses the philosophy behind this feature in more detail.

The audio windows provide little more than an indication that the audio track is active when it is visible. Research has been carried out on incorporating audio analysis facilities into *Video Transcriptor*. This consists of graphical representations of the peak and average audio levels for each audio track in the movie being analysed. This graphical representation has been aimed to assist in determining when speech begins and ends by the subject. The features examined have not been incorporated into *Video*

Transcriptor. Section 8.6 — *Audio Support* in Chapter 8 discusses this idea in more detail.

6.4 Transcription Windows

The main concept of a digital video analysis system is to assist in the analysing and the note-taking involved. The transcription windows are the primary source for note-taking under *Video Transcriptor*. When a new researcher is confronted by *Video Transcriptor*, they are only able to open one transcription window to hold their notes. As they become more familiar and use the more advanced modes that are available, they may decide to have separate transcription windows for each track in the movie.

The transcription window is constructed from three main components (figure 6.5). The time resolution shows the movie time, the current time is highlighted in red. The event category component allows the user to define groups for notes. The third area contains the details about individual events.

The time resolution component allows the user to change the time interval that the events are displayed at by double-clicking the mouse in this area. For example, instead of being every 0.08 seconds, it can be changed to be every two seconds (figure 6.5). More information about the time resolution of movies can be found in Section 4 — *Standard Mode Features* of the *Video Transcriptor User Manual* under the heading — *Movie Time Resolution*.

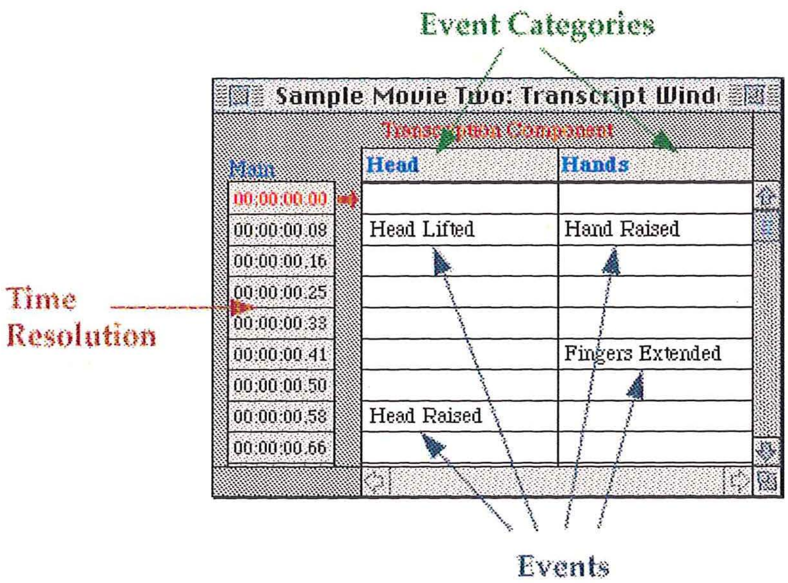


Figure 6.5. Transcription Window.

The concept behind the design of the transcription window has been to mimic the idea of notes that are written using pen and paper techniques. This whole concept is very similar to setting up columns on a piece of paper that had previously been used in the analysis of video tape. Figure 6.5 shows an example for the analysis of human behaviour. Two features of human behaviour are to be analysed based on the subjects head and hands. These are to be the event categories that our individual events should be entered into.

The user begins the transcription phase by first opening a transcription window. The researcher must then define the categories that are to be examined, for example in figure 6.5 there are two categories defined, *Head* and *Hands*. Many of the functions provided in the note-taking process involve the use of pop-up menus. These are used for creating and modifying the event categories and all functions involving event entries. The pop-up menu is activated by pressing the mouse button down and holding it down while pointing in either the event category or the event areas. Pop-up menus provide speed of use instead of having to use pull-down menus all of the time. The functions are not sped up by having keyboard short-cuts since individual event categories or events have to be initially selected with the mouse anyway. If the entry is selected then it is inefficient to then drag the mouse up to a menu item or to release ones hands from the mouse to select a keyboard short-cut. Keyboard short-cuts are not practical since the user would have to remember about a dozen extra keys.

Entering new events can be done by using the pop-up menu available by clicking and holding the mouse button down while pointing at a specific event location. Double-clicking in the same location can also have the same effect in certain conditions. By selecting *New Event* from the pop-up menu, a window for entering the event details appears (figure 6.6). This is the central component to the note-taking facilities that *Video Transcriptor* offers. The timing information is available at the top of the window while the user enters the details about the event below this.

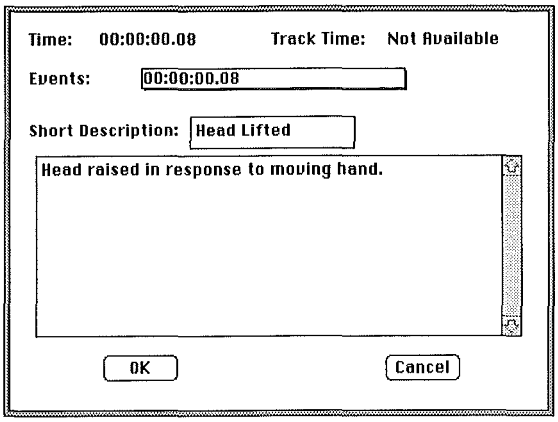


Figure 6.6. Event Entry Window.

Initially this may seem like a slow process compared with pen and paper techniques, but it does have its benefits. It is not necessary to go through this process every time. As the user enters new events, a list of all the different types of events are built up into a list. For example, in figure 6.5 the short description "Head Lifted" is added to the list of unique event types. Each time the event pop-up menu is activated to enter a new event, a list of all these event types appear at the top of the list (figure 6.7). By selecting an entry from this list, an event entry is automatically inserted into the transcription window. This saves considerable amount of time when there may only be a couple of different event types that are entered in each event category. This is a common feature when analysing video information. For example, the researcher may be interested in looking at the head movement of the subject involved and there may be only six possible ways for the head to move. As the user enters the details about the head, a unique list is built up offering all the different event types that have occurred for a specific event category. The user only needs to select the entry from the pop-up menu to automatically create a new event entry. It is faster than writing down the same information, once the event is initially entered into the window in figure 6.6. The entries are located at the top of the pop-up menu so that they can be accessed quickly when entering notes. Figure 6.7 shows an example of a pop-up menu.

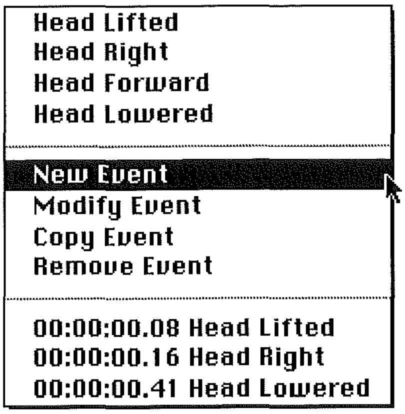


Figure 6.7. Sample Event Pop-up Menu.

The middle section of the pop-up menu offers standard editing functions that can be performed on the events, and the bottom of the list contains a list of all events that have occurred within the same time resolution. The order of these items once in the pop-up menu relate to the frequency that they are used. More information about the note-taking process can be found in Appendix A — *Video Transcriptor User Manual.*

Video Transcriptor's transcription windows only support frequency based event types. It does not support duration based transcriptions due to the nature of research being undertaken in the psychology department. The reason for not including this support is due to the testing carried out did not involve this type of analysis.

6.5 Opening New Movies and Exporting Transcripts

Video Transcriptor does not offer the capability of recording new movies or making edits to existing movies. These capabilities are found in many QuickTime compatible applications, many of which are free. It makes no sense in duplicating other people's efforts when these facilities are already freely available.

However, *Video Transcriptor* offers the capability to open additional movies in order to add tracks from these movies into the current movie that is being analysed. This capability is needed in order to analyse multiple video and/or audio tracks simultaneously. Analysis of multiple tracks is a feature that was needed by the psychology research students. They often have to try to analyse the same subject expressing different emotions based on the same material they are being subjected to, or try to analyse different subjects. This is a very slow process when using a single video tape player and pen and paper. This type of accurate comparison has not been popular since the facilities are very rarely available. With the ability to analyse multiple tracks simultaneously, more accurate results are achieved and will open up new types of experiments that can be easily compared against each other.

Video Transcriptor does not offer any analysis facilities. There are many statistical analysis packages available to perform these complex facilities. The easiest way to supply compatibility from *Video Transcriptor* to these statistical analysis packages is by exporting the desired results out to a text file, whereby it may then be imported into a statistical analysis package. Exporting the notes taken within *Video Transcriptor* is available by selecting *Export Transcript* from the *Transcription* menu. The resulting text file contains a text format of the transcript. Development of this feature has not been considered a high priority since the emphasis is on the interaction with the system and the benefits that it has to offer. Exporting the information is a straight forward task. The information is readily available but the output format needs tailoring to the specific package requiring it. Facilities for the user to choose the order and delimiters of fields need to be made available for the user to create customisable export files to be compatible with any statistical analysis packages available.

Chapter 7

Evaluation

To evaluate the effectiveness of *Video Transcriptor*, users were assessed on several factors. These included:

- the time taken to learn the system,
- the difference in productivity with and without *Video Transcriptor*, and
- user' feedback to the interface components.

The users involved in the evaluation of *Video Transcriptor* included a research student from the Psychology Department, computer science masters students, and friends. All of the users provided feedback on interface issues, while the Psychology student provided responses to the note-taking facilities.

7.1 Descriptions of Tests and Evaluation Methods

Initial design stages involved finding out what the researchers were currently doing and trying to find solutions to the limitations and problems they face.

The design of the system was prepared on paper as a small user guide, showing exactly what it would provide. This was shown to the Psychology student who made suggestions for changes to the proposal about the basic principles of the design and interface. After coming to an agreement on the initial design, work began on a prototype system, know as *Video Transcriptor*. This was developed for Apple Macintosh computers.

Once the prototype system was developed, it was shown to the Psychology student to evaluate the interface and contents. The session involved the research student attempting to use the system based on what they had read from the design document, and direct interaction through the interface they were confronted with. The subject's actions were videoed throughout the session to provide feedback to the areas where problems existed. Interface problems with several components on the Master Controller were revealed through this process. During this session the user found it difficult to grasp the concept of several of the main buttons for controlling the video — the control for setting the start and end times of the active movie portion, and several other elements in the interface. Changes were needed to these areas of the controls. This led to the introduction of user skill levels into the main controller, whereby the user can progressively learn about the interface in stages, without becoming overwhelmed with all of the controls at once The interface problems encountered are discussed in Section 7.3.

Problems encountered were fixed and suggestions from users incorporated into *Video Transcriptor* and then tested again. Testing and evaluation of *Video Transcriptor* has occurred throughout the development stages. Feedback from users has led to many interface changes. Figure 7.1 illustrates the life-cycle of the typical interaction with users.

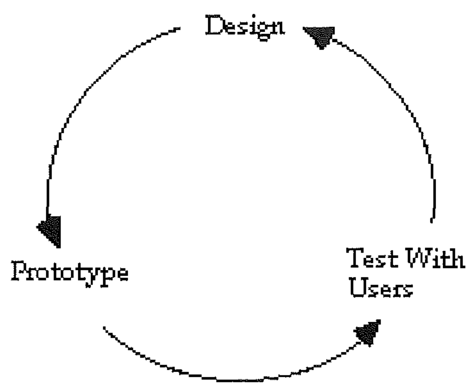


Figure 7.1. Testing Life-Cycle With Users.

Subsequent tests involved observing the research student using the system without recording the interaction onto video tape. The information needed was obtained from observing the user and discussions about the interface.

Changes were made to the basic shape of the Master Controller from a horizontal window across the bottom third of the screen, to a vertical window located against the left hand edge of the screen. The reason behind this decision was to make a stronger analogy with the concept of a remote control unit for a video tape player. Consequently the buttons on the Master Controller look and behave similar to a video tape player's remote control. The users felt more comfortable with this shape since they could easily relate it to their personal experience using a remote control on a video tape player. Users were more willing to use the buttons for controlling the video without the hesitation they had when the Master Controller was at the bottom of the screen.

Observations of the research student evaluating *Video Transcriptor* showed a significant improvement in time for analysing video information compared to video tape methods. Feedback about the quality of the controls offered by *Video Transcriptor* was positive — the control of the video information was vastly improved in the eyes of the researcher.

Evaluation of the present video tape methods employed by the research students, showed that they were very time consuming. This has lead many researchers who were interested in using video tape to avoid it due to the excessive amount of time involved.

In addition to the testing with the Psychology research student, and feedback from other masters students, an experiment was carried out with a friend to determine the difficulties presented with conventional video tape analysis systems and *Video Transcriptor*.

The person involved had no previous experience with analysing video information. The reason for the choice of person was so they would have no bias or preconceptions. If the person is unaware of what is involved, then they can approach both methods from an equally independent viewpoint. The aim of the test was to make a comparison between conventional note-taking and a digital video based system.

A short section of video material was chosen from Oliver Stone's mini-series "Wild Palms" recently shown in New Zealand. The video section to be analysed by the user was one minute and fifty seconds in length and consisted of a male and female character talking to each other. The chosen section of video tape was also digitised onto a computer, taking about twenty minutes to compress, for the user to analyse with *Video Transcriptor*.

The aim for the user was to locate all the times that a specific character closed their mouth. The shots of the subjects in the program were mainly of their waist and head.

Initially the user analysed the section on video tape using pen and paper to write down the exact time that the event occurred. The user was told to take as long as needed to make sure that the times taken were as accurate as possible.

Once the user had finished analysing the video tape section, they required a period of time to learn how to use *Video Transcriptor*. Initially the user explored the system by themselves and then was given the section from the user manual labelled "Quick Start", refer to Appendix A — *Video Transcriptor User Manual*. This section of the manual is intended to guide the user through a sample session. Once the user felt confident with what they had learnt, an additional feature was taught. This consisted of using the pop-up menu to assist in automatically entering events needed during the testing phase.

The user required less than fifteen minutes to learn about the system. At the end of this time they were capable of performing all of the required features needed for using *Video Transcriptor* including some of the more advanced features of the note-taking process. This was comparable to the time other people had taken to learn about the system when they were analysing and testing it.

7.2 Comparisons Against Hand Techniques

The user took fourteen minutes to analyse the video tape information with the accuracy of the information recorded only to the nearest second. Typically in research facilities the video tape would have timing information recorded in one corner allowing a higher precision in the accuracy of results. With the extra timing information, the researcher also spends significantly more time in analysing the information to get the exact time. In the test carried out, the user observed the subject closing their mouth twenty two times during the time period.

In comparison, using *Video Transcriptor* took twelve minutes to analyse the video information. The difference in time between the two methods is negligible from these results. There were several factors to consider when looking at these results. The first involves the user's inexperience with both methods used. Pen and paper techniques were natural to the user so little learning was needed. *Video Transcriptor* however was new, even with a short training session. The speed of the user was significantly slower than other people who had been involved in the interface testing over a period of time. Consequently, the speed of the user would be expected to improve as a result of more usage.

Another consideration to the results was the poor quality of the compressed video in some sections in the digitised video that had to be examined. The user spent the majority of their time trying to accurately analyse these sections. As a result, their time effectively doubled in comparison to other sections of the video that they analysed.

The user found thirty events in the section of video when using *Video Transcriptor*. This is eight more than using video tape showing the higher level of accuracy available. Video tape was significantly harder to analyse. Each time the user missed the time of an event they had to rewind the tape and try again. *Video Transcriptor* allowed the user to step back and forward frames until they got the exact location, quickly and easily. Statistically the results are not representative of what might be achieved due to the small number of subjects involved in the test.

7.3 Interface Evaluation

There were many interface issues that needed addressing. The main one was changing the shape of the Master Controller so it closely resembled a video tape player's remote control for reasons already discussed.

Video Transcriptor contained too many visual elements on the Master Controller, often distracting and confusing a new user. This lead to the implementation of three skill levels of operation; basic, standard, and expert, each progressively displaying more of the features on the Master Controller. Other redundant information was also removed during this process, including statistical information about the movie like the number of frames per second, and early versions of audio controls.

Initial versions of the Master Controller contained three rows of buttons for controlling different aspects of the movie, which was reduced to two rows. The bottom row of five buttons originally contained two volume buttons; one for increasing, the other for decreasing the volume; and a third button for a mute. These were changed to allow finer control of the audio level in the movie. The new volume control is based on a slider control which contains more visual feedback for the user.

The last two buttons on the last row were the *record*, and *repeat* buttons. The record button was removed since *Video Transcriptor* does not support the recording of movies because there are already many free programs to do this. The *repeat* button was relocated, replacing the *still* button that had the same function as the *stop* button.

The note-taking facilities underwent many changes over the testing period. Initially the transcription window was attached to the bottom of the video and audio windows. This restricted the user from resizing the windows. Initially note-taking consisted of double-clicking in the cells of the transcription window to create and modify events. Pop-up menus were introduced to assist in note-taking by providing the automatic insertion of new events based on existing entries, and the removal of unwanted events. This did not provide enough facilities for manipulating the information. Subsequently, additional menu items were introduced to alleviate these problems. The new menu items added included copy and paste features to improve the environment.

The method of entering notes in *Video Transcriptor* has been the source of initial confusion for many users. They are faced with the transcription window and have been able to successfully create new event categories from the menu item *New Event Category*. However, they can find nothing similar for the entry of events. Those people who read the section in the user manual found out what to do easily, while others worked it out by exploring. Some people however are not willing to test out the interface for fear of breaking something that they are unable to fix.

7.4 Analysis of Video Transcriptor

Discussions with the users indicated that their preferences tended towards *Video Transcriptor* as a more efficient means for controlling the information and taking notes. The main reason was the fast and highly accurate controls that were available when compared to video tape technology. Other factors included the speed at which the user was able to enter event details after finding the exact location of the event.

The only real disadvantage was the image quality being poor in places for trying to analyse the details of the video. It was difficult to make out the shape of the subject's mouth since they were a distance from the camera at certain points in the video segment. When the shots on the video were close up there were no problems at all. Improvements could have been made by trying other compression methods for the video to improve the quality, or adjusting the various controls available for hue, saturation, brightness, contrast, sharpness, black, and white levels. The number of controls available for adjusting the picture quality makes it difficult for the lay person to get the "perfect" picture. Typically, the default settings produce a high quality result without any need for adjustment. There is the potential for losing some image quality but this depends on the source material, compression techniques chosen, and the image adjustments made. The quality of the required information is dependent on what it is to be used for — analysing body movements does not require a high quality picture when compared to analysing a subject's eyes. The quality will improve as digital video gets better with the introduction of bigger and faster disks, better capture techniques, and improved compression methods.

Discussions after the user's interaction with *Video Transcriptor* highlighted the effectiveness of the note-taking. When analysing the video tape, the user was required to write down the timing information each time. In contrast, entering the details was automatic in *Video Transcriptor,* simply the selection of an item in a pop-up menu. The similarity of controlling the video was another feature noted, making the transition easier for the user.

Even though the use of *Video Transcriptor* is not initially any faster than using video tape methods, the speed does increase the more it is used. Furthermore, the quality of the notes taken is higher than those using pen and paper with video tape. Events that had previously passed unnoticed using video tape are now more likely to be picked up due to the improved control of the information. Additionally, the times of recorded events are significantly more accurate.

Chapter 8

Implementation Details

Video Transcriptor is the result of researching previous and current time based analysis systems, the evaluation of various user interface designs, and the assessment of user's interactions with the system.

This chapter discusses the many issues that have arisen in the development life-cycle of *Video Transcriptor*. These issues include the choice of development platform, language structures, the intricate details that had to be overcome once development work had commenced, and features under development.

8.1 Platform For Development

The choice of the development platform had been decided through the analysis of the different digital video systems available. Special emphasis was placed on the cost and accessibility of the platform requirements. The chosen platform was Apple Macintosh computers using QuickTime. Apple Computer's QuickTime system architecture had been chosen for several reasons; the software component was free, it was an integral part of the operating system and not an add-on feature like many other schemes, and it was available on other platforms for future support. Other significant factors in this choice included the supply of machines within the Computer Science Department, and more importantly the Psychology Department where the software is to be used. Macintosh computers are relatively common within educational institutes which makes it an ideal platform choice. The developed system does not require the purchase of any new computers or software but is able to utilise existing facilities. In the current scenario this works well with the Psychology

Department who mainly use Apple Macintosh LC computers. The cost is a very important issue to consider. Since Universities and other research facilities do not always have surplus finances to spend. The use of Apple Macintosh computers with QuickTime make it an ideal platform that incurs no costs to initially set-up.

8.2 Language Used

The development environment chosen for *Video Transcriptor* is Symantec's THINK C. There are several factors behind this choice of language and compiler. The language C is one of the most widely supported development languages on the Macintosh so it is the logical choice to use. The other main development languages used are C++ and Pascal but both to a lesser extent. Other supporting reasons for choosing C is the experience from previous development work and the accessibility of resources within the department.

Macintosh Programmer's Workshop (MPW) C is an alternative compiler to Symantec's THINK C for the Macintosh but the cost, availability, and immediate support for it locally has made it a less attractive alternative. There is a general tendency towards using Symantec's development products in recent years mainly due to the relatively cheap cost of their development tools and the integrated environment provided.

C was chosen as a language not only because of the support it has as a development environment but also because of its speed and versatility. The requirements for a graphical based system of this nature is that it demands extensive processing power for handling digital video information in real-time. This involves the extraction of the digital information from the storage media and the decompression in real-time. The speed of playback and user response is important. True object oriented languages such as Smalltalk are inappropriate for this type of system. They typically require as much memory as they can get and they can potentially slow the system down at random times while garbage collection occurs. This situation is unsuitable when a user is expecting reasonable feedback to actions at all times. Interrupting *Video Transcriptor* while playing a movie in order to reorganise the system memory is simply not acceptable for users.

8.3 Structure of Program

The development of *Video Transcriptor* on Apple Macintosh computers has involved a steep learning curve. Initially coming to grips with an event-driven, graphical user environment takes time to adjust to when coming from a totally procedural programming background. The steep learning curve is a standard feature whenever programming in a new environment. Once new or difficult features of the development have been overcome and mastered, subsequent efforts in doing similar things become significantly easier.

Limitations in the knowledge of the Macintosh programming environment initially lead to many non-optimal design decisions. Many of these have been reworked and improved from the initial design and testing stages.

The current version of *Video Transcriptor* has been developed using Symantec's THINK C compiler making use of the QuickTime Toolbox, which provides standard functions for handling time based information. Not all the features required are supported by the QuickTime Toolbox. Several have been implemented via alternative methods. These features include multiple video displays and audio analysis support which are discussed later in this chapter.

Video Transcriptor has been a constantly changing system as it incorporates new features and encompasses modifications and improvements to existing features as design and interface issues are tested. Several major structural changes have taken place due to improvements being made in QuickTime. One such change had been an alternative method for displaying multiple video windows simultaneously. *Video Transcriptor* allows for any number of movies to be loaded, each of which may have any number of tracks. The only restriction imposed is the amount of memory available. Settings are preserved for an individual's workspace each time *Video Transcriptor* is used. The window location of the Master Controller and other global pieces of information are stored in a preference file when the user quits. The transcription notes taken are stored within the movie file. The notes taken in the main transcription window are stored within the movie's user data area while notes taken in individual track's transcription window are stored in the track's user data area. This isolation of the notes allows for the tracks to be moved around into different movie files, preserving the transcription notes.

8.4 Human-Computer Interface

"There are no guidelines regarding ways to allow users to set the rate of a movie; ..."

- Guillermo Ortiz [35].

This is an example of the type of problems facing the design of a digital video analysis system. The whole concept of using digital video as an analysis tool is relatively new and the interface needs careful consideration. Funding and resources are not available to launch into major research into the best approach available, but reasonable assumptions have been made based on previous popular metaphors and users responses. Larger corporations like Apple will have their Human-Computer Interface teams draw up recommended interface designs for people to use after extensive research. Wendy Mackay's system, EVA, was built upon the Athena Muse multimedia platform that took 8 years and $100 million to develop, [6, 31]. Controlling the rate at which the movie is played back poses an important interface problem. Should the user have a push button for pre-set speeds, a slider style control for adjusting the speed, or a shuttle control like that found on popular Mitsubishi video-tape recorders. There may be more types of controls to consider — a power-glove attached to the system, speech recognition, and possibly more. How do we decide on what is best? Recommended standards that developers should follow are good for keeping applications consistent, but what does one do when there are no such guidelines in place and who says that the recommended guidelines are the best approach? What one person may find easy may be totally against the natural way of doing things for another user. This is a major consideration that has been faced in the development of *Video Transcriptor*. Interface designs were trialed and subsequently changed as users tested and suggested alternative methods, or they found the interface components too confusing and difficult to use.

An easy-to-use graphical user interface requires careful consideration and involvement with people who are to use the system. It is the users who ultimately end up using it day in and day out, not the programmers of the system. They have valuable information that needs to be extracted.

The interface design can be broken up into two main areas; the controls for the digital video elements, and the note-taking process. The design of these two areas has been based on existing ideas. Initial and subsequent design and testing stages have been described in Chapter 5 — *The Design of a Digital Time Based Analysis System.*

One facility that was required by the psychology research students was the capability to examine aspects of the video information in greater detail. Standard video tape players [6, 31] do not offer any capability for this. *Video Transcriptor* provides an interface for examining the video based information in greater detail. This is in the form of a control that allows the user to resize a viewing box around the required section of the original source material. The area within the viewing box is resized to fill the entire window, enlarging the video contents. The viewing box is the portion of the whole image that is always seen in the video windows. More information about the function of the Zoom Controller can be found later in this chapter and in Appendix A — *Video Transcriptor User Manual.*

Psychology students often have to make comparisons between subjects, or even the same subject at different times as part of their analysis of human behaviour. *Video Transcriptor* offers the capability of having multiple video tracks visible simultaneously making comparisons between subjects easy. Comparing the same subject at different times normally involves searching the video tape for the required segments. This can take several minutes while the tape is physically being searched. The implementation details behind this are discussed in Section 8.5.

The second main aspect of the interface design is the note-taking facilities that are provided within *Video Transcriptor*. Essential elements behind the design of the note-taking is in the ease of use and the power of the facilities provided.

Many of the refinements to the interface elements have been based on user experiences. This has applied not only to the note-taking but also on the general controls available. The users have preconceptions about how the interface should operate. The method of operation should reflect similar attributes of their current system. This is an important aspect that has been incorporated into the design. The design of elements in the interface has involved subjecting the user to a variety of interface concepts to determine their reactions. From the involvement with the user in experimenting with many different controls the good and the bad aspects can be determined.

8.5 Multiple Track Support

Another feature required by researchers is the ability to examine multiple subjects simultaneously. The QuickTime system architecture provides support for storing multiple video and audio tracks, within a movie file. However, the first releases of QuickTime did not provide any means for displaying multiple video tracks simultaneously into different windows, and even the later versions only provide limited means for this. Multiple audio tracks were, however, fully supported.

Two separate methods were used to support the playback of video tracks simultaneously, the first involving multiple movie files synchronised by their time bases, and the second uses new functions supported by QuickTime 1.6.

8.5.1 Synchronised Time Bases

The initial versions of QuickTime, (beta, v1.0, and v1.5), did not support any capability for the simultaneous playback of video tracks into different windows. Any number of audio tracks could be played back but the video tracks could only be displayed into a single window.

These versions of QuickTime had an internal limitation of only allowing one graphics window to be assigned as the display window for all video tracks to use. It is possible to split this window up into smaller sections and draw the contents of the different videos into these smaller areas but this is limiting and unnatural for the user who should be able to arrange them how they would like.

Simultaneous playback of multiple video tracks can be achieved by loading additional instances of the movie for each video track in the original movie, and then synchronising them. Audio tracks can be used from the first movie file since they do not require being displayed into a single window. Figures 8.1 and 8.2 illustrate the concept of loading multiple movie files for each video track.

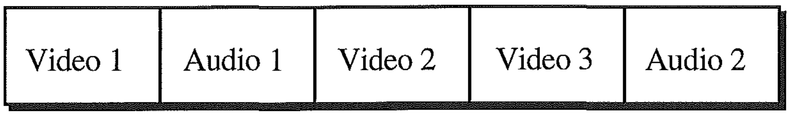


Figure 8.1. Original Movie File Containing 3 Video and 2 Audio Tracks.

Figure 8.1 illustrates the construction of the original movie. It has a total of five tracks, three video and two audio. What is required is for two more instances of the original movie file to be loaded for the second and third video tracks. The first instance of the movie supplies the first video track and the two audio tracks while each subsequent instance only has one active video track that is used. Figure 8.2 illustrates the original movie and the additional copies of the original movie that must also be loaded.

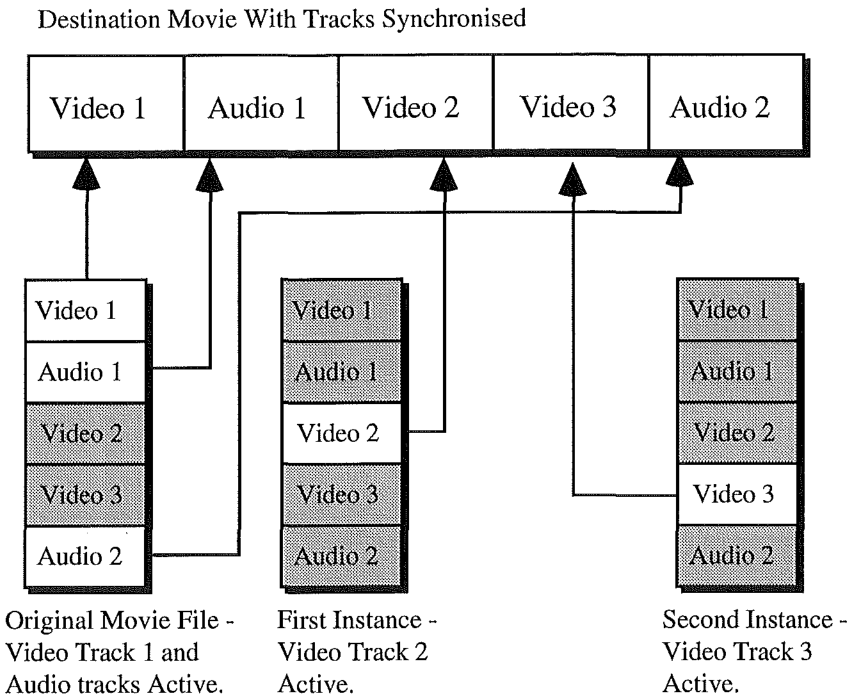


Figure 8.2. Loading Multiple Instances of the Original Movie.

This method is at the expense of system memory. For example, opening a sample movie file took 180KBytes of memory and each subsequent time the movie had to be opened took another 50KBytes. Each time the movie was reopened it used information already available from the first movie to save memory.

Once the multiple copies of the original movie are loaded there is still the problem of synchronising them. When two or more instances of a movie have to be played back simultaneously they would normally behave independently of one another. This was solved by enslaving all of the additional movie files time bases to the original movie, as suggested by Guillermo Ortiz [35]. The concept behind this is to have control of the movie's time flow pass through a single point instead of having a number of individual time bases running at the same time. By forcing this time base of the original movie to be used by the other instances only one movie needs to be controlled. What this effectively means is that by changing the time in the original movie, all the enslaved instances of this movie will automatically update their own current time to this new time. The structure of *Video Transcriptor* had to be rewritten to accommodate multiple movie files all being linked together. All major operations had to be linked back to the original movie file so that only one central copy of the movie was being modified. The secondary copies of the movie file were only to be used as containers for displaying the extra video tracks required.

8.5.2 Support For Multiple Video Tracks in QT1.6

QuickTime 1.6 gives some support for simultaneously displaying multiple video tracks. It is not a complete solution, but offers many advantages over the previous solution of enslaving the time base of several instances of the same movie.

The new support enabled a video track to draw its contents into a specific graphics window, (normally an off-screen area of memory), which may be different from the one specified for the whole movie. After the track contents have been drawn they are normally copied to the actual movie's graphic window. This solution improves *Video Transcriptor's* memory intensive nature. Instead of drawing the track's contents into the off-screen graphics areas and then copying them to the movie window a better solution is to reverse this scenario. By forcing each video track to write directly to a separate window and making the movie window effectively an off-screen area of memory we can get the desired results. This facility was not intended to be used this way but supports what is required.

This addition of assigning independent graphics windows to tracks in QuickTime 1.6 had a major effect on the structure of *Video Transcriptor*. Major rework and fine tuning was needed once again to change the underlying structures to incorporate this enhancement. The main changes was how the video tracks were handled. Instead of loading multiple copies of the same movie and synchronising them, one movie was to be loaded and different graphics environments needed to be set up and maintained.

Other subsequent changes involved the resizing of video windows. Previously it was an independent action of the other video tracks since they were copies of the same movie. Each movie containing a single active video track would automatically know the size of the current movie's video display area. Resizing a video window would only change one movie's display area independently of the other movie files. Under QuickTime 1.6, any changes made to a single video track's window size results in wider consequences. Changing the size of a single video window affects the movie's display region, which is the basis for all video windows. Thus all video tracks are dependent on the one global size at any time.

One simple approach, that is not desirable, would be not allowing the user to resize the size of the video tracks. To examine the video information in close detail the user would only be able to use the zoom control feature on the master controller.

Another method — the one that has been implemented — is to allow the user to resize the video window to any size desired and then have the system resize all other video windows within the movie to the same size.

A method that works, but not implemented in the final design due to time constraints, allows individual resizing of video windows. This involves transformation matrices for each video window allowing the video in each track to be resized to the current video window's size independent of the other video windows. This would take the size of the movie's display area and apply a scaling factor based on each video window to resize the contents of the video. More detail on this principle is covered later in this chapter in Section 8.7 — *Zoom Controller.*

8.6 Audio Support

There is currently little support for audio analysis in QuickTime. Researchers would find it helpful to have the ability to analyse the structure of the audio tracks to determine when sounds begin and end with a visual display. QuickTime only provides the playback of the sound and nothing more.

A similar visual representation to an audio level meter or a graphic/spectrum equaliser on Hi-Fi equipment is needed when trying to find the start and end points of human speech. By looking at the wave forms of the sound pattern it is easy to determine when these events occur. Typical displays available for this type of analysis consist of peaks, average levels of audio during discrete time intervals, and real-time representations of the sound pattern as it occurs.

Research has been undertaken into providing better support in this area. When the movie file is loaded a copy of the audio tracks are made into separate Macintosh sound handles. This process requires additional memory for each audio track. We do not require loading additional movie files so the additional memory saved previously can be redistributed.

The sound handles that are extracted from the movie file conform to the standard Macintosh 'snd' resource format for sounds. This contains header information at the beginning and then the raw sound information. The format of the information includes the frequency of samples and whether the audio information is compressed.

The main problem is the amount of information that is available in the audio tracks. Typical sound may be sampled at about 11, 22, 44.1, or 48kHz. The actual information stored in a 22kHz sound consists of about 22254 sound samples per second. This is too much data to be processed in real time to obtain average information or peaks on top of the current video and audio decompression already taking place. This is especially true when there may be several audio tracks to process simultaneously. To calculate accurate information for each second of audio information in a single track takes more time than it does to play the sound. This means that it is unrealistic to perform any real-time analysis that uses all of the information available.

A better approach, at the expense of memory usage, is to calculate the peak and average values of the audio tracks when the movie file is first loaded. By creating two arrays for each audio track, the average and peak levels for each frame in the movie can be pre-calculated and stored for instant access. If a one minute movie is taken that contains 25 frames per second, and each sound value can be recorded into a single byte the memory required is:

> 25 frames per second x 60 seconds x 1 byte per frame
> = 1,500 bytes per minute

for storing each of the average and peak levels of the audio. For a typical session of three minutes, the requirements are

> 1,500 bytes per minute x 2 arrays (peak and average)
> x 3 minutes
> = 9,000 bytes for a single audio track.

The memory required is minimal even for several audio tracks. One hour of audio information would require 180,000 bytes of memory. Typical sessions are only several minutes long with several audio tracks. The amount of memory required for a typical session is minimal, offers accurate results, and can be pre-calculated. 180KBytes of memory is not an excessive amount of memory by current system configuration standards, but the main problem is the time that is needed to pre-calculate these results.

Another method that can be used to assist in processing this information is a probabilistic analysis approach. One feasible method involves taking every *n* samples out of the sound information. This would effectively reduce the rate at which the sound was sampled. For example, if we took every five samples from a sound recorded at 20kHz it would approximately reduce it to a 4kHz sample. This can reduce the number of samples to be processed in real time but is not totally accurate. A true 4kHz sample would average the five values available.

If we had a movie with five video and audio tracks then the overhead in processing this amount of information can be too much. A more appropriate scheme would be to take a very small number of samples and use them to calculate an estimated audio value. Typical analysis used by research students is only required to find out the start, end, and volume of sound. For example, if ten samples were taken from 100 samples of sound, we could calculate the probability, P(x), that one of these results was in the top twenty percent of the samples in order to estimate the peak level of the audio.

By using the formula

$$P(x) = 1 - \prod_{i=0}^{m-1} \left(\frac{k - i}{n - i} \right)$$

where

n = population,
m = number of samples randomly chosen, and
$k = n \times q$ = target population, where q = target percentage.

we can calculate the probability that one of the twenty samples is in the top twenty percent of sample.

After putting the appropriate entries into the formula above we get a probability of 70.26% that one of the ten samples will be in the top twenty percent of the values. By increasing the number of samples that we choose from, to twenty, we get a 99.90% probability that one of the twenty samples will be in the top twenty percent of the values. By just doubling the number of samples we are almost guaranteed that one of the samples will be in the desired range.

In real-life scenarios the sample values of audio information are not evenly distributed and are very dependent on previous sample. Nevertheless, even though this approach is only a very rough estimate, it can produce a reasonably accurate representation of the audio without having to over-extend the system requirements.

8.7 Zoom Controller

Implementation of the Zoom Controller required a series of matrix operations being applied to the video content of the movie through calls within QuickTime.

There were two completely different structural approaches to the Zoom Controller. This was due to structural changes being applied from supporting multiple video tracks (Section 8.5). The initial controller was written when multiple copies of the same movie were loaded for each video track within a movie file. The current control is based on only one movie file being loaded.

The initial design was based upon multiple copies of the same movie loaded all linked together through their time bases. The user resizes the controller to determine the area that is to be expanded in the video windows. The size and location of the control determines the area that is to be expanded. The video is then resized proportionately by using matrix operations and then translated so that the top left hand corner of the window will have the same visible area as the top left corner of the control area. Once these two operations have been done the final image is clipped to fit into the video window.

Transformation matrices are used to define graphical operations on the movies while they are being displayed. A transformation matrix defines how to map points from one co-ordinate space into another. By modifying the contents of a transformation matrix, graphical operations including translation, rotation and scaling can be performed.

QuickTime uses a 3-by-3 matrix to perform these two-dimensional operations (figure 8.3).

$$\begin{bmatrix} a & b & u \\ c & d & v \\ t_x & t_y & w \end{bmatrix}$$

Figure 8.3. Structure of the 3-by-3 transformation matrix.

It assumes that the matrix elements u and v are always 0 and the value of the matrix element w is always 1.

The 3-by-3 matrix transforms a point (x, y) into a point (x', y') by means of the following equations:

$$x' = ax + cy + t_x$$
$$y' = bx + dy + t_y$$

The identity matrix, that performs no transformations, is shown in figure 8.4.

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Figure 8.4. The Identity Matrix.

When a user resizes the Zoom Controller the video image has to be resized to reflect this operation. As the zoom area rectangle gets smaller the area of the video images have to be scaled by a larger factor. The horizontal and vertical lengths of this resizable rectangle over the size of the controller are used as the scaling factor necessary to resize the images (figure 8.5).

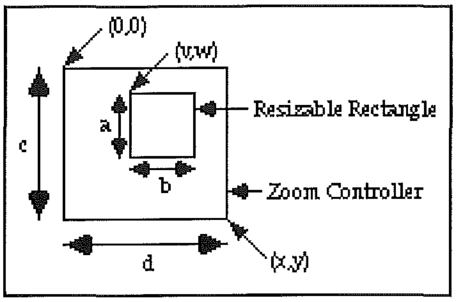


Figure 8.5. Scaling Factor.

From this we can see that the horizontal aspect of the video image has to be increased by: $\frac{d}{b}$x100%, and the vertical aspect by: $\frac{c}{a}$x100%.

This is achieved by using a scaling transformation matrix to modify the x and y co-ordinates by $S_x = \frac{d}{b}$, and $S_y = \frac{c}{a}$ (figure 8.6).

$$\begin{bmatrix} S_x & 0 & 0 \\ 0 & S_y & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Figure 8.6. Scaling Transformation Matrix.

Once the video image has been resized, the image has to be shifted so that the contents within the area of the resizable contents are displayed in the video windows at the correct position. This can be done by using a translation matrix to shift the x and y co-ordinates (figure 8.7), where $t_x = -v,\ t_y = -w$.

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ t_x & t_y & 1 \end{bmatrix}$$

Figure 8.7. Translation Transformation Matrix.

These two operations can be combined by combining them into a new matrix by matrix multiplication (figure 8.8).

$$\begin{bmatrix} S_x & 0 & 0 \\ 0 & S_y & 0 \\ 0 & 0 & 1 \end{bmatrix} \times \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ t_x & t_y & 1 \end{bmatrix} = \begin{bmatrix} S_x & 0 & 0 \\ 0 & S_y & 0 \\ t_x & t_y & 1 \end{bmatrix}$$

Figure 8.8. Combined Matrix Transformations.

An example of the transformations required for resizing the zoom area are shown below. Figure 8.9 shows the area on the Zoom Controller that is to be examined in detail.

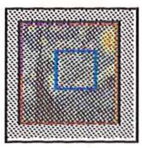


Figure 8.9. Zoom Controller.

The image is scaled in the x and y directions by the amount previously calculated. This will scale the zoom area of the image so that it now fills the window (figure 8.10).

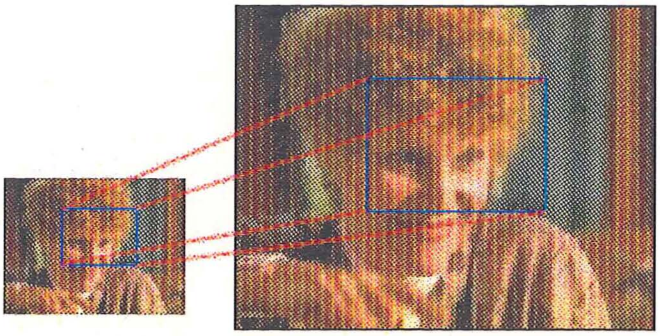


Figure 8.10. Scaled Image.

The boxed area in the right hand image is the same size as the initial window containing the entire image. This is the area that we need visible in the window. At this stage the top left hand corner of the new image is visible so a translation is required to see the contents of the zoom area (figure 8.11).

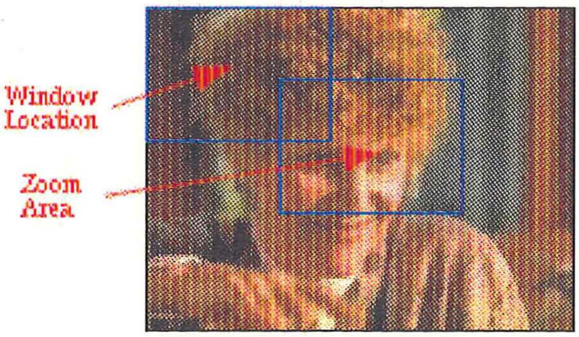


Figure 8.11. Translation of the image.

The final process is to clip the viewing area to the required window co-ordinates (figure 8.12).

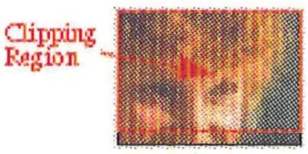


Figure 8.12. Clipping Region of the Video Window.

The final result is the video being scaled, translated, and finally clipped to fit into the required area in the video window, figure 8.13.

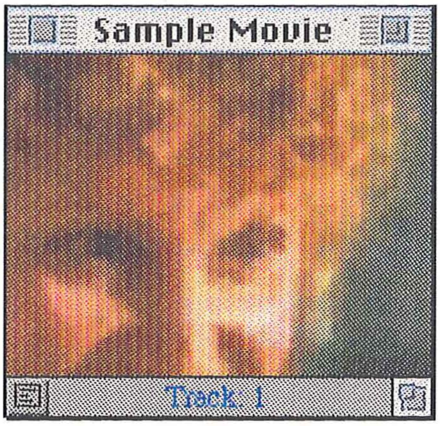


Figure 8.13. Final Image After Scaling, Translation, and Clipping.

Under QuickTime 1.6, features relating to clipping the display area of a movie are no longer effective since the level of control had changed to an individual track level. Clipping of the movie's display region only applies to one video window that the movie is normally displayed in. Under the previous method each video track was loaded from a separate movie file with the video window also acting as the movie's display region.

In QuickTime 1.6 each individual video track is loaded from the same movie file and displayed in a separate window with the movie's display area effectively being unused. Functions that affect the movie's display area like clipping no longer work since there is no longer a movie display area. Typically each track region would be clipped to their respective sizes and then would all be brought together into the one movie display window which could then be further clipped and processed. This however does not work, using individual display windows for each video track. Support for clipping regions on an individual track level are supported but these only apply globally to the movie's display region and not at an individual track's level. The clipping region that is applied to the track does not affect the individual window that it is drawn into. It affects the track if it had been drawn into the movie's display region. After the tracks' clipping regions are all calculated then a movie clipping region could be applied. This movie clipping region was what was previously used.

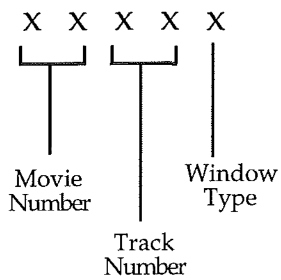
The standard Macintosh Toolbox routines for clipping a windows display region work, but QuickTime overwrites this clipping region when it displays the video image data into the window. It effectively overwrites the standard Macintosh Toolbox routines, replacing them with only those supplied in the QuickTime Toolbox routines.

This clipping of each track's display region is required to correctly display the image data into a window. The bottom of the window contains a button for the transcription button in the bottom left hand corner, information about the track number in the centre, and the grow box in the lower right hand corner. Without the support for clipping regions for individual tracks in QuickTime 1.6 there is no immediate solution. This information at the bottom of the window is overdrawn when the movie is playing or updated. To resolve this issue, the information is redrawn whenever the video information is stationary.

8.8 Window Control

An issue in the implementation of developing a system containing many types of windows and movies is controlling and knowing what type of windows are displayed, the movie each belongs to, and the tracks they are associated with. Windows on the Macintosh are generic, and additional information is required to be stored with each window data structure to hold this information.

The information that has been stored in *Video Transcriptor* is a number in the following format,



where the *Window Type* is either a video, audio, transcription, or the Master Controller window. The *Track Number* is the track that the window is associated with in a movie and the *Movie Number* is the movie that the window is associated with.

8.9 Note-taking Facilities

The note-taking facilities that were developed and incorporated into *Video Transcriptor* support discrete time intervals for events for frequency based note-taking. It would be possible to provide support for duration based note-taking in the future but the emphasis is on finding out how beneficial the system is for researchers.

The note-taking in *Video Transcriptor* is handled in the *Transcription* windows. There is one window available for notes about the entire movie, and individual windows available for each track in the movie in the more advanced levels of the program.

Initial designs of *Video Transcriptor* had the transcription facilities attached to the bottom of the video and audio windows when it was activated. This design was changed for two reasons. The first was when a large video window was being used and the transcription facility was activated, the video window doubled in size to accommodate the note-taking facilities at the bottom of the screen and subsequently became inaccessible. The second reason for moving this was the fact that it was very difficult to arrange the windows on the desktop. By making the transcription facility a separate entity on the desktop it still allows the windows to be played together and is more flexible.

Activating the individual transcription windows from the menu bar is a context sensitive operation. This follows Apple's noun/verb metaphor used in the desktop. When a user selects an object like a window, icon, or text, they then use the menu to perform an action to the particular object. For example, when a user selects some text they can choose several operations that can be performed on that piece of text like *cut*, *copy*, and *paste*. The same applies to activating transcription windows in *Video Transcriptor*. The user selects a particular track by clicking in the window of that track, or clicking on the Master Controller if the main transcription window is required, and then they select *New Transcript* from the Transcription menu.

Event details can be entered automatically from a pop-up menu in the transcription window. This speeds up the note-taking process significantly when there are large numbers of similar types of entries.

Copying events is also made simple through the pop-up menu together with changing the contents of existing events. Events can be moved around as easily as copying and pasting text in a word processor.

The note-taking information is stored with each individual track in the movie, or, alternatively with the movie file itself. The movie, and each track, support a *User Data* area for placing information into. By storing the notes with the actual movie and track data the user's transcription notes will not get lost. The notes are stored in the user data areas in a delimited format when the movie data is saved. When a movie is loaded, the information is parsed into internal linked list data structures within *Video Transcriptor* where it is modified by the user during note-taking.

The structure for the user data information that is stored with the movie is as follows

User Data:	Track Time Resolution :: [Event Category List] <<
Event Category List:	Event Category ' ' [Event Details] ;; [Event Category List]
Event Details:	Time :: Short Description :: [Long Description] :: [Event Details]

where entries in "[ ]" are optional. If one of the delimiter characters are found within a field entry, a '\' character is inserted before it (when saved), or removed (when loaded), so it can handle delimiter strings within the text.

The internal data structure for the transcription notes in memory is illustrated below as a series of linked lists (figure 8.14).

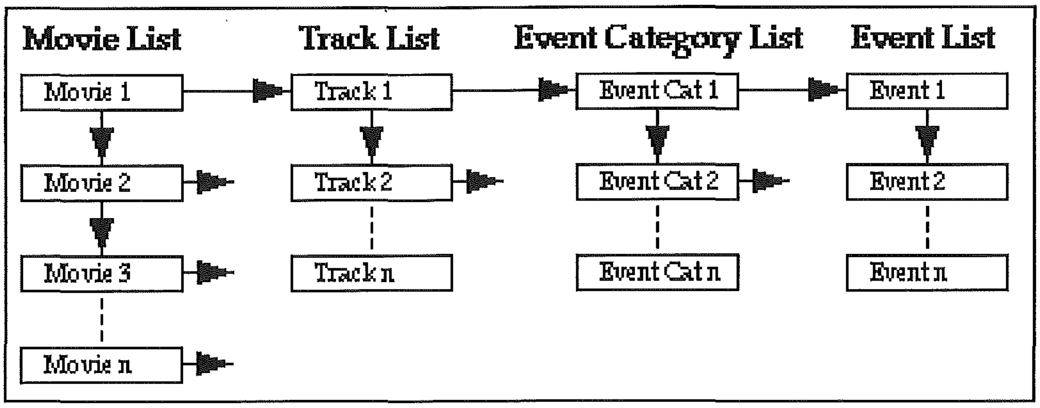


Figure 8.14. Internal Linked List Structure For Transcription Notes.

Video Transcriptor can have any number of movies open at once, each containing a series of tracks. Each track in the movie contains a series of user-definable event categories. Within each event category is a series of events.

8.10 Variable Time Resolution

Video Transcriptor offers the researcher the ability to change the default time resolution of the note-taking process for each track, or movie. By default, the interval displayed for entering events is the time it takes to display one frame of video information. This may be too fine for some purposes.

Video Transcriptor allows users to change the resolution by clicking in the *time resolution* component on the transcription window. As a result, this changes the basic functions for the majority of the note-taking facilities. When a new event is entered, or an event is to be modified, within a time interval, the user must select the time for the event within the time interval. All notes are still entered with an exact time value. The transcription window either displays the event details in a time interval, if there is only one event, or the number of events that occur within the time interval.

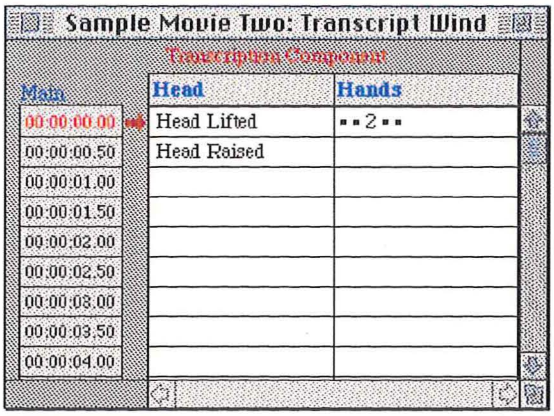


Figure 8.15. Modified Time Resolution.

For example, in figure 8.15, the time resolution is half a second. The first two time intervals in the "Head" event category each have a single event entry. The first event entry in the "Hands" event category has two events entered, as indicated by the entry "• • 2 • •".

More details about the consequences of changing the time resolution can be found in Appendix A — *Video Transcriptor User Manual,* Section 4 — Standard Mode, under "Movie Time Resolution".

8.11 Exporting The Results

The emphasis of the research work has been on the interaction and effects that digital video can have on time based analysis methods. The export feature is very basic at present. The exported file consists of a readable text file containing the transcription notes.

More advanced formats for the exported data can easily be written. A better way to provide flexibility is to produce a totally configurable format editor for the exported data. The format would need to provide input into the many statistical analysis packages already written.

8.12 Features Under Development

There are several facilities that are currently under development. These facilities are discussed in depth in Section 5 in Appendix A — *Video Transcriptor User Manual* at the end of this thesis. The facilities that are covered include:

- Audio analysis within the audio windows.
- Audio control on the Master Controller.
- Tools for manipulating individual tracks.
- Speech track support.
- More advanced exporting facilities.
- The ability to offset new tracks that have been appended to the movie file.

Most of these are enhancements relate to the analysis of the audio tracks.

This topic of the analysis of audio information has been discussed in Section 8.6. It is a useful facility for in the analysis of audio information found in the movie files. To coincide with the extra audio development, there also needs to be additional controls available for setting individual volume levels for each track.

Tools to assist in the analysis of individual tracks are also required. All of the controls currently available apply to all tracks in the movie. Individual tools for analysing each track independently of the others for comparison purposes would be useful to complete the set of available note-taking facilities.

With the emphasis on audio information being examined, a speech track facility requires development work. This has two aspects to it; to assist in the annotation of speech, and the ability for the user to make speech annotations. The first scheme offers the facility for recording what users are saying during any interaction. The second scheme allows the researcher to attach recorded messages into the note-taking facilities to improve the resources available in the analysis of video and audio information.

The current exporting features available in *Video Transcriptor* are somewhat limited. They only offer a formatted text file in an easily read format. User configurable exporting features are needed so *Video Transcriptor* can interact with other statistical analysis packages. This is considered a relatively minor task to produce the desired results.

With the ability of adding additional tracks to a movie file for comparison purposes the matter of synchronising the information needs attention. This consists of changing the offsets that each track starts from the beginning of the movie file until they are all aligned together.

Chapter 9

Conclusions

This thesis provides a study into the benefits that digital video can bring to time based analysis techniques. The most important result is the improvement in the analysis stage that researchers obtained from using digital video.

Digital video is a powerful medium offering a great potential in assisting with the analysis of time based analysis systems. The quality, speed of playback, and disk space required are the main problems that have faced digital video. Full-screen, full-motion, broadcast quality video requires additional hardware assistance to handle the extra information needed, and large hard drives to store all of the information. QuickTime is an entirely software based solution offering adequate performance for no cost.

The three main problems that face digital video are no longer a barrier. Improved video compression schemes and hardware provide larger, higher quality, video at faster frame rates. Likewise the storage requirements are also reduced. For the majority of analysis work, the speed and quality of the compression schemes available with QuickTime are sufficient. Other considerations are the processing power of personal computers, increasing steadily each year, and storage devices are getting larger and faster. Consequently the cost of the processing power and storage devices are continually dropping, and, the access to digital video is easier and cheaper than ever. Machines like the Apple Macintosh Quadra 840 AV and the Centris 660 AV, released by Apple in August 1993, offer video input and output as standard features allowing for the capture of digital video from any source. Likewise Silicon Graphic's low end "Indy" workstation, released around the same time, comes bundled with a digital video camera to sit on top of the monitor. These are just the beginning of personal machines aimed at supporting digital video in the everyday work environment.

9.1 New Capabilities

Digital video provides the capability to develop new features that are not possible using video tape technology. These features have opened the way to better and more accurate analysis of the information provided.

Video tape is a very time consuming medium to work with. Typical times for analysing a single behaviour in three minutes of information range from forty minutes to one hour. The majority of the time spent is trying to reposition the video tape. Digital video systems offer instant access and much finer control over the video information. The use of *Video Transcriptor* showed significant improvements in the speed and accuracy achieved.

The ability to compare multiple subjects proved to be very valuable to the research students. The research students reported that comparison based results were more accurate, the speed of note-taking was dramatically improved, and the results were based from the primary source of information. Results are often derived by comparing the secondary source of information, (the notes taken by the researcher), instead of relying on the primary source of information, (the original video tape source). The main reason for this is the time required to compare two subjects directly from video tape. With *Video Transcriptor* the comparisons were more accurate since all of the subjects are visible simultaneously, while only one is possible with video tape.

Detailed analysis of certain visual areas of subjects has been accomplished through the ability to resize the video contents. Research students need to analyse certain features of body movement in detail such as eye movement. Video tape offers no easy solution for examining information closely. Research into eye movement is extremely difficult using conventional video tape technology. By enlarging the area around the subject's eyes, the analysis was significantly easier for the researcher. The original image quality is vital for this to be of benefit for the researcher. The better the original video material digitised, the better the results that can be obtained. A clean video source means better compression and little degradation of the video image, making it easy for the researcher to analyse.

Increased speed in the analysis and note-taking of single behaviours was observed through tests performed. The increased speed was by researchers using *Video*

Transcriptor compared to using video tape with pen and paper for the note-taking. When analysing a single behaviour there may be several possible outcomes to be noted. *Video Transcriptor* "learns" about the notes that the user has taken previously. Each time the user has to enter a new event, they are offered the possibility of selecting a previously entered event to insert. In contrast, hand-written notes require the user to re-enter the notes everytime. The tests showed the increased productivity by easily inserting the notes after finding the exact location of the video information.

The incorporation of the notes and the video system into one package lead to improved note-taking methods. Pen and paper methods often resulted in misplaced results. With *Video Transcriptor,* the results are stored with the video information so they are never lost or misplaced. This direct relationship with the video information makes analysis significantly easier as a result.

The design of the system is such that it can be used to teach other researchers how to analyse video information. Each researcher can take down notes about events that they perceive to happen and then they can all be compared to check the accuracy in their observations.

Overall control of the interface is initially straight forward for users. Elements are extracted from their previous knowledge, like the interface for controlling the video information. Elements of this have been derived from a remote control for video tape players. Users exhibited enough knowledge to open movies and control their playback without any knowledge of the system. As the users spent more time with the system, and explored the features, the more they learnt. The response of some controls was of concern on the slower machines like the Macintosh LC. Experience with a faster machine illustrated that it was merely a problem of the slower machine having difficulty handling the vast amount of information required. Machines are continually getting faster and cheaper. The processing power found in the high end machines of Macintosh line several years ago are now being shipped as entry level machines.

9.2 Problems

The main problem facing digital video is the disk requirement needed to store the information. Disk space is a relatively expensive commodity when compared to alternatives like video tape. It is not normally available as a long term storage media. In Chapter 3 alternative storage devices were compared. Large hard drives are preferable for intermediate storage for speed purposes but alternative storage media such as CD-ROM disks and tape units are more affordable as long term storage.

These alternative options have their own limitations — CD-ROM media is read-only, and tape media is sequential making access slow. Despite these limitations both media options offer all of the advantages that digital video have to offer over analog video tape methods.

Other problems facing digital video technology is the time required to digitise the video and then compress it. In the Psychology Department, research involving video tape normally involves recording half hour sessions of each subject. For each subject, a three minute section is chosen and then recorded onto a separate video tape. Instead of this duplication process, whereby the image quality severely deteriorates after each generation, the section of video can be digitised instead. This process of digitising the video takes the same amount of time that copying it takes. However, to save significant disk space the information needs to be compressed. This created several problems; the time to compress the information, and the resulting quality. The one minute and fifty second section of video and audio information used for experiments took about twenty minutes to compress. The compression of the digital video significantly reduces the amount of storage space required but it can also reduce the quality of the original video information, depending on the compression scheme used and the settings applied. As a result the image quality may be less than ideal for analysis purposes. This was the case for the test video that was recorded and used in experiments.

The sample video recorded was not the highest quality and subsequently when the user was required to analyse certain portions of it to find out when the subject's mouth closed, the quality of the video was somewhat poor. This is a problem to be aware of when digitising and compressing video information. It is important to have a good quality source of video and take care in selecting compression options.

Features that need addressing include the analysis of the audio information. Time constraints have limited development in this area. When analysing video tape in slow motion, no audio information is available. Digital video allows audio playback when playing the information back at slow speed but not when the researcher typically steps through the information. Representation of audio information is needed to assist in finding certain events in the video information.

Digital video proves to be an efficient means for the analysis of traditional video tape sources. Increased control of the video information and integrated transcription facilities has lead to faster note-taking and improved accuracy in researchers results. The system itself can be run on any colour capable Macintosh computer without having to purchase additional software. The only requirement is access to a video digitising card which may be borrowed or is relatively inexpensive to purchase. Digital video makes a realistic alternative to video tape analysis methods.

Bibliography

1 Mark Anbinder, Personal Communication, New York (October 19 1992).

2 Joseph Ansanelli and Doug Camplejohn, *QuickTime and the Competition*. Version 1.0, WorldWide System Software Product Marketing (1991).

3 Allan Bell, Personal Communication, Developer Services, Australia (10 December 1993).

4 Don Crabb, Multimedia Products for the Macintosh and Windows. *Higher Education Product Companion,* **2** (3): p22-26, 1993.

5 Tony Crease, Putting Multimedia to Work. *MacNews*, p22-26, July 1993.

6 Alan Dix, Janet Finlay, Gregory Abowd and Russel Beale. *Human-Computer Interaction*, Prentice Hall, (1993).

7 John W Donovan, Intel/IBM's Audio-Video Kernel. *Byte*, p177-186, December 1991.

8 Starkey Jr Duncan and Richard E Sayre, C-QUAL: A System For Computer-aided Transcription of Videotaped Action Sequences. *Behavior Research Methods, Instruments, & Computers,* **23** (4): p505-512, 1991.

9 Michael Floyd, Data Compression and the AVK. *Dr Dobb's Journal*, p18, July 1992.

10 William Fulco, The QuickTime/AVK Connection, Building a Beautiful Relationship. *Dr Dobb's Journal*, p28-30, July 1992.

11 R P Futrelle and M J Potel, The System Design For GALATEA, an Interactive Real-time Graphics System For Movie and Video Analysis. *Computers & Graphics,* **1** p115-121, 1975.

12 Didier Le Gall, MPEG: A Video Compression Standard for Multimedia Applications. *Communications of the ACM,* **34** (4): p46 - 58, 1991.

13 Clifton Lee Gass, A Digital Encoder For Field Recording of Behavioural, Temporal, and Spatial Information in Directly Computer-accessible Form. *Behavior Research Methods & Instrumentation,* **9** (1): p5-11, 1977.

14 James L Green, Capturing Digital Video Using DVI, Multimedia and the i750 Video Processor. *Dr Dobb's Journal,* p16-24, July 1992.

15 Joint Photographic Experts Group, *JPEG (ISO/CCITT) Still Image Compression Standard.* 5th Edition, (1989).

16 Moshe Guthertz and Tiffany Field, Lap Computer or On-Line Coding and Data Analysis For Laboratory and Field Observations. *Infant Behavior and Development,* **12** p305-319, 1989.

17 Kevin Harney, Mike Keith, Gary Lavelle, Lawrence D Ryan and Daniel J Stark, The i750 Video Processor: A total Multimedia Solution. *Communications of the ACM,* **34** (4): p65-75, 1991.

18 Hugh Honour and John Fleming. *A World History of Art,* PaperMac, (1982).

19 Apple Computer Inc. *Apple Announces Cross-Platform Directions.* 1991.

20 Apple Computer Inc. *Apple's QuickTime Begins Shipping.* 1991.

21 Apple Computer Inc. *The Apple CD-ROM Handbook, a guide to planning, creating, and producing a CD-ROM,* Addison-Wesley Publishing Company, (1992).

22 Apple Computer Inc. *QuickTime Developer's Kit, version 1.5,* (1992).

23 Apple Computer Inc. *QuickTime Technical Backgrounder.* 1992.

24 ISO, *Coding of Moving Pictures and Associated Audio.* ISO11172: ISO/MPEG 90/176 Committee Draft, (1990).

25 Robert M Krauss, Palmer Morrel-Samuels and Jonathon Hochberg, VIDEOLOGGER: A Computerized Multichannel Event Recorder For Analyzing Videotapes. *Behavior Research Methods, Instruments, & Computers,* **20** (1): p37-40, 1988.

26 Wolfgang Kreutzer and Bruce McKenzie. *Programming For Artificial Intelligence,* Addison-Wesley, (1989).

27 Brenda Laurel. *The Art of Human-Computer Interface Design,* Addison-Wesley Publishing Company, (1990).

28 Brenda Laurel. *Computers as Theatre,* Addison-Wesley Publishing Company, (1991).

29 Michael Liebhold and Eric M Hoffert, Toward an Open Environment for Digital Video. *Communications of the ACM,* **34** (4): p103-112, 1991.

30 Andrew Lippman, Feature Sets For Interactive Images. *Communications of the ACM,* **34** (4): p92-101, 1991.

31 Wendy E Mackay, EVA: An Experimental Video Annotator for Symbolic Analysis of Video Data. *SIGCHI Bulletin,* **21** (2): p68-71, 1989.

32 Brian Michon, Integrating Motion Video into Computational Environments. *SIGCHI Bulletin,* **21** (2): p80-82, 1989.

33 Richard F Olivo and Margaret C Thompson, Monitoring Animal's Movements Using Digitized Video Images. *Behavior Research Methods, Instruments, & Computers,* **20** (5): p485-490, 1988.

34 Kip Olson, Making Better QuickTime Movies. *Develop - The Apple Technical Journal,* p93-106, June 1993.

35 Guillermo A Ortiz, Time Bases: The Heartbeat of QuickTime. *Develop - The Apple Technical Journal,* p41-52, December 1992.

36 W B Pennebaker and J L Mitchell. *JPEG: Still Image Data Compression Standard,* Van Nostrand Reinhold, New York, (1993).

37 Alan C Repp, Martin L Harman, David Felce, Rick Van Acker and Kathryn G Karsh, *Conducting Behavioural Assessments on Computer-Collected Data.* Northern Illinois University, Southhampton University Medical School, and British Institute of Mental Handicap

38 Les Paul Robley, Digital Offline Video Editing: Expanding Creative Horizons. *American Cinematographer,* p68-72, 1993.

39 Jeremy Roschelle and Shelley Goldman, VideoNoter: A Productivity Tool For Video Data Analysis. *Behavior Research Methods, Instruments, & Computers,* **23** (2): p219-224, 1991.

40 Neil B Sharman, *An Empirical Comparison of Progressive Image Compression Techniques*. University of Canterbury (1992).

41 Frans Sijstermans and Jan van der Meer, CD-I Full Motion Video Encoding on a Parallel Computer. *Communications of the ACM,* **34** (4): p81-91, 1991.

42 Harold Thimbleby. *User Interface Design,* Addison-Wesley Publishing Company, (1990).

43 Maarten J van der Vlugt, Menno R Kruk, Annemoon M. M Van Erp and Reint H Geuze, CAMERA: A System For Fast and Reliable Acquisition of Multiple Ethological Records. *Behavior Research Methods, Instruments, & Computers,* **24** (2): p147-149, 1992.

44 Gerald M Walker, Cable Television: Preparing a Compressed Future. *World Braodcast News,* p48-51, 1993.

45 Gregory K Wallace, The JPEG Still Picture Compression Standard. *Communications of the ACM,* **34** (4): p30-44, 1991.

46 Gordon Werner, Personal Communication, (February 10 1994).

Appendix A

Video Transcriptor User Manual

Contents

Preface

Video Transcriptor is a software-based note-taking tool for use in the analysis of video and audio time-based information. This manual explains how to install and use *Video Transcriptor* on the Macintosh computer. It is a tutorial for learning the basic concepts and progressively the more advanced features of *Video Transcriptor* and the note-taking procedures available. The manual is divided into the following sections:

Quick Start

A simple guide to getting started immediately. This provides a brief look at what *Video Transcriptor* can do for you.

Section 1 Introduction

A description of what *Video Transcriptor* offers and the advantages over more conventional video-tape technology.

Section 2 Installing *Video Transcriptor*

An explanation of the system requirements and how to install *Video Transcriptor.*

Section 3 Tutorial

An example of how to use *Video Transcriptor.* This covers many of the basic issues faced by the end user in the note-taking process.

Section 4 Standard Mode Features

This section covers additional features available when in the STANDARD mode of *Video Transcriptor.*

Section 5 Expert Mode Features

This section covers advanced features available when in the EXPERT mode of *Video Transcriptor.*

The manual begins with a brief introduction to using *Video Transcriptor* in the form of a **Quick Start** guide.

The manual concludes with a glossary and an index.

Quick Start

This section is designed to give you a quick taste of *Video Transcriptor*. It is not intended to be a comprehensive introduction but a brief overview. For a more detailed tutorial refer to Section 3 - *Tutorial*.

In this quick overview you will start the program, open a video, control the playback of a video, analyse and take notes on a video, and export the results.

After installing the software[1], start the program by opening the *Video Transcriptor* folder and double-click the *Video Transcriptor* program icon. This will bring up the Master Controller as shown in figure 1.

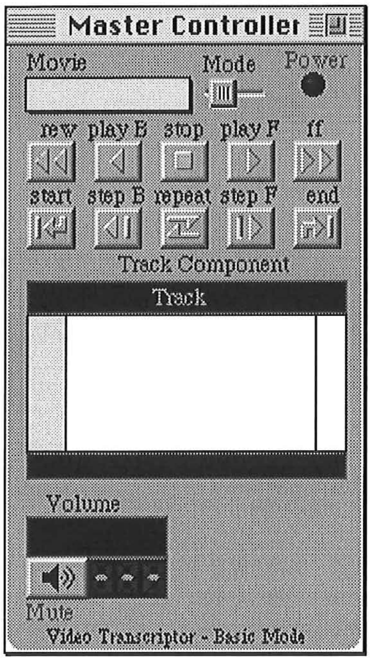


Figure 1. The Master Controller Window.

[1] For installing *Video Transcriptor* refer to Section 2 - *Installing Video Transcriptor*.

Opening a Video

To begin using *Video Transcriptor* to analyse and take notes on video and audio data, a QuickTime format **movie** must be loaded. *Video Transcriptor* does not create QuickTime format movies. These will normally be created with the assistance of additional hardware. In this quick reference the "Sample Movie" supplied has already been created for you to practice with.

To open the movie, select the *Open Movie* command from the File Menu. Figure 2 shows the file selector box that appears.

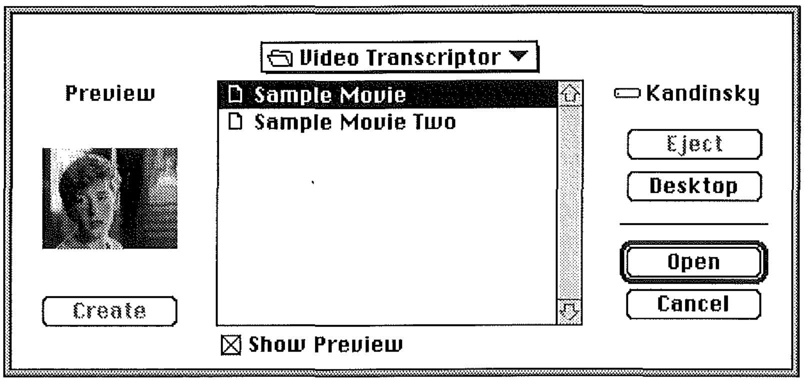


Figure 2. The Standard Macintosh File Selector Box.

Select the "Sample Movie" and press the *Open* button. This will load the movie ready to be analysed.

Playback the Movie

Once the movie has been loaded it can be viewed by using a set of buttons that resemble a video recorder's remote control. These buttons are located at the top of the Master Controller. Figure 3 shows this set of basic controls that are available.

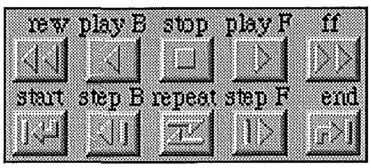


Figure 3. The Playback Controls.

Press the "play F" button to start playing the movie and then stop the movie by pressing the "stop" button. Press the "play F" button again and watch it play to the end of the movie and stop. When the movie has stopped at the end try pressing the "start" button to make it return to the beginning of the movie. Keep practising with these controls until you are comfortable with their operation.

Once you are familiar with these controls try pressing the "end" button to take you to the end of the movie. Press the "play B" button to play the movie in the reverse direction and watch it stop at the beginning of the movie.

These buttons covered allow you to control the movie in a similar manner that you would control a video in your video player at home.

Try pressing the "repeat" button. The button changes shape into a circular arrow shape. Press the "play F" button and watch what happens when the movie reaches the end. You will notice that it starts over again from the beginning, looping forever until it is stopped by using the "stop" button. To turn off this looping feature just press the "repeat" button again.

Other controls that are found on a video player are the "ff" and "rewind" buttons. Try these two buttons to fast forward and rewind the movie at a high speed. Use the "stop" button to stop it at any point while searching through the video.

Press the "start" button to return to the start of the movie. Start playing the movie by pressing the "play F" button and then press the "stop" button somewhere in the middle. The movie can be analysed frame by frame by pressing the "step F" and "step B" buttons. These buttons will move the movie to the adjacent frame in the movie in either direction.

These buttons should assist in analysing the contents of the "Sample Movie" in detail. When you are finished exploring the buttons for controlling the movie, press the "start" button before going onto the "Transcription Process".

The Transcription Process.

Figure 4 shows an example of a transcription window with notes.

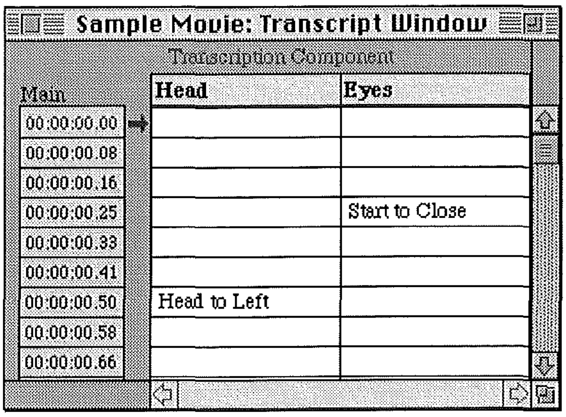


Figure 4. Transcription Window Containing Notes.

We will now create the same notes as we can see in figure 4. Select the *New Transcript,* command from the Transcript Menu. An empty window for entering notes will appear as in figure 5 below.

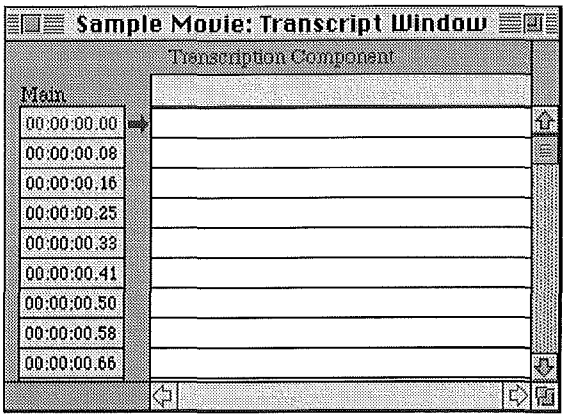


Figure 5. The Transcription Window.

First we will define the two categories that our notes are to be about; "Head" and "Eyes". To create these categories, select *New Event Category* from the Transcript Menu. Figure 6 shows the window that appears to enter the category name.

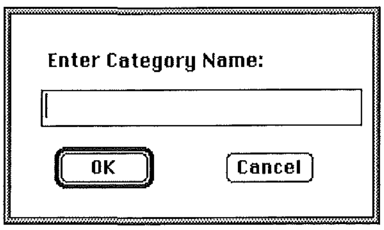


Figure 6. Dialog Box For Entering Categories.

In this window enter "Head". Repeat this process of creating a new category except enter "Eyes" the second time instead of "Head" for the name of the category. Two categories should now exist in the transcription window for notes to be entered.

The transcription window should now look like figure 7.

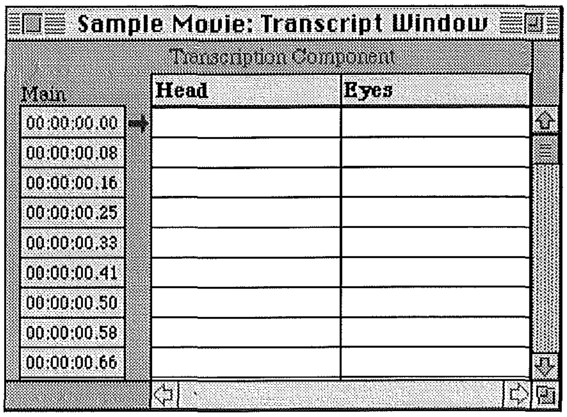


Figure 7. Transcription Window With Two Event Categories Defined.

We can see that the movie is at the beginning, from the current time at 00:00:00.00. Try using the "step F" button on the Master Controller to step through the movie frame by frame until you reach the time "00:00:00.25". This is the point at which we will take our first note about eye movement. The transcription window should have the time "00:00:00.25" highlighted in red with an arrow next to it.

To enter our first note about eye movement, double-click the mouse in the column labelled "Eyes", beside the arrow pointing at time "00:00:00.25".

A window appears, figure 8, ready for information about the event to be entered.

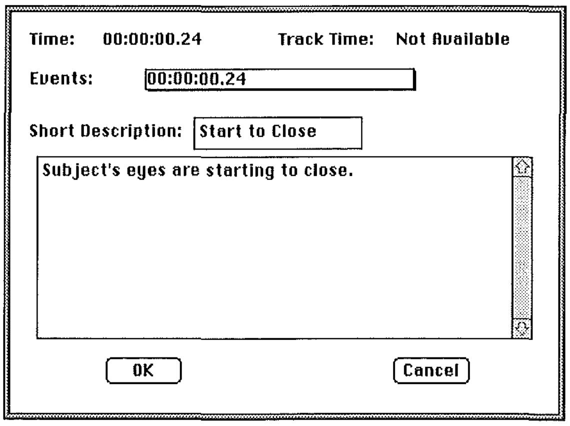


Figure 8. Event Dialog Box.

Enter "Start to Close" in the short description field. This will be shown in the transcription window when all the information is entered. Enter "Subject's eyes are starting to close." into the larger field below the short description field. This larger field holds a detailed description about the event. When finished entering the information, press the "OK" button to save the event data.

The transcription window now shows "Start to Close" in the "Eyes" column at time "00:00:00.25".

Repeat this process for the second event shown in figure 4 at time "00:00:00.50". This time it is an entry about the subject's head movement so it will go in the column labelled "Head". Double-click in the column "Head" at time "00:00:00:50". This will once again bring up a window for you to enter some information about the event. This time enter what is in figure 9 below.

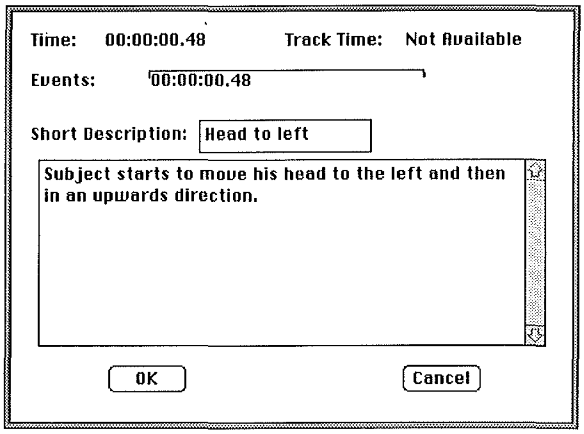


Figure 9. Event Dialog Box.

When you are finished press the *OK* button.

To change any information about one of these events, just double-click on the event entry again. This will bring up the same window that was used to enter the information initially. This shows the basic principles for entering the information into *Video Transcriptor*. The next step is to export the results of what you have just entered.

Exporting the Results

Select *Export Transcript* from the Transcript Menu. This will ask for the file name that the exported results will be written to. Press the "Save" button when you are ready to save the transcription data. To view the exported data, double-click on the saved file.

This will load it into the application "TeachText" for viewing. Refer to Section 5 - *Expert Mode Features* for more details on exporting the results of the note-taking process.

Section 1

Introduction

What is Video Transcriptor?

Video Transcriptor is a software based note-taking tool for use in the analysis of time-based data such as plant growth, human behaviour and high speed physics experiments. It is oriented towards the analysis of video and audio data. *Video Transcriptor* is not designed for a specific user-group in mind. It is aimed at people who wish to analyse any type of time-based information. The advantage of *Video Transcriptor* over other systems is the total integration of the time-based information with the note-taking procedures on a single computer. The most common analysis method used by researchers consists of using a video-tape player to hold the time-based information and paper to store all the notes. This cumbersome process can now be made easier with the assistance of *Video Transcriptor*. By recording all of the video-tape information onto the computer, the researcher may examine the information in greater detail and accuracy than with a video-tape player. All of the note-taking is also done on the same computer in a user-configurable environment for the researcher.

The ideal system for the analysis of video or other time-based information is one that the user feels they have total control over the entire system. This means very precise controls for moving through the data at any speed and direction and the ability to skip to any location instantly without the wait that conventional video tape systems inherit while the tape rewinds.

Another desirable feature is the ability to integrate the note-taking with the actual data so instead of having to write the resulting analysis on paper and then associate it with a video tape, both can be combined. By doing so, the real-time data and the associated user's notes can be distributed as a whole. This is a significant advance in having the note-taking as an integral part on the computer. A major advantage in having the note-taking facilities incorporated on the computer is the automation of repetitive tasks. One such repetitive task is in the note-taking process where the same information has to be written down all of the time. This is easily taken care of by trying to reduce the intervention needed by the user, thus leaving more time for other purposes.

Many systems have been developed in the past containing facilities for computer-aided note-taking but they all still rely on video-tape as being the primary media on which the time-based information sequences reside. These systems are restricted by the speed of the video-tape player for accessing the physical tape. Another consideration is the expense for all the electronic components which make up these systems for transcription analysis of videotape sequences. In the case of many of these systems the cost may outweigh the benefit of the system for users with a limited budget.

The idea of using a computer instead of a video-tape player is to let the computer help the user in the analysis of real-time data. The system can be as easy or as complex as the user requires, and thus no limitations imposed upon the user at the outset. The computer itself can be used for other general purposes like word processing. The only equipment necessary is access to a video digitising card that is only used briefly to capture the raw data to be analysed.

With the computer there comes the power of instant and accurate access to any position in the movie. Even with the power of fast, slow and still display of the tracks, there is no physical deterioration of the media upon which the movie is written, unlike video-tape media. Other features available with a computers assistance is the automation of time consuming transcription operations include being able to select from a common pool of transcription event labels instead of writing the same item down repeatedly.

Brief Overview

Video Transcriptor is different from existing products. *Video Transcriptor* draws on elements from other programs together with features desired by end-users for analysis. All of these elements have been drawn together into one integrated environment to assist in the analysis of time-based information.

Only recently has it been possible to integrate the video and audio elements of the source material directly onto the computer. Video-tape was the primary source for analysing this type of data. Apple Computer recently released a standard extension to their Macintosh based system software, known as QuickTime, which has made this important step possible. Not only has it allowed the integration of video and audio data onto the computer, but it has changed the way people interact with computers. The advantages of the new system extension is that it is distributed free of charge and will work on "any" colour capable Macintosh computer without any additional hardware or software. Computers other than the Macintosh require the purchase of additional hardware to produce similar results, which may be expensive and can lead to incompatibilities from the proprietary formats of the data captured. QuickTime is now available on most platforms, making it easier to share movie data.

Video Transcriptor takes a previously digitised **movie** as its source. The content of this movie can be analysed, by using the features of *Video Transcriptor*. Tools are available to step through the movie at any desired speed and to analyse any segment of the movie in close detail. These tools are accompanied by a large number of note-taking facilities.

On completion of the note-taking process, the data can be exported into flat text files from which it can be imported into popular statistical analysis packages, databases and other software systems.

What are the Pitfalls of Current Video-Tape Technology?

Video-tape information has always been a laborious and difficult exercise to analyse in the past. The researcher is normally armed with two tools; the video-tape player's remote control unit, and a pen and numerous pages of paper. Many hours are spent repeatedly going over and over the same segment of video-tape while the researcher frantically takes notes about any events that occur. This entire process is not only slow but also awkward for the researcher. Organising the hand-written notes with the video-tape is not always an easy task. Finding the appropriate section of notes with a segment of video information is even harder, especially when it is needed in a hurry. The video-tape technology available is not the most accurate available. Searching for a segment of video information can take several minutes to find while the tape is being wound. The researcher is limited to the available controls of the video-tape player in order to closely examine any information on the video-tape. Controls normally available include normal playback, fast playback (shuttle search), step forward and backwards, and a pause button (depending on how good it really is).

It is difficult to make comparisons between subjects directly. Usually the two subjects will reside on different video-tapes or different locations on the same video-tape. Comparison of the two subjects usually has a distinctive pause between locating them either while the video-tape is being swapped or being wound to its new location. With a second video-tape player this process can be made easier.

Improved control of the video information is possible by using a video editing suite. These usually offer reasonably accurate control of the video but the expense of purchasing the extra and more expensive equipment may outweigh the benefits of the improved accuracy.

What are the Benefits of Using *Video Transcriptor*?

Video Transcriptor offers many benefits over traditional video-tape based systems. Some of the pitfalls of video-tape technology have already been discussed. Some of the features that *Video Transcriptor* has to offer include:

- an integrated environment for both the analysis of the video information and the note-taking process.
- full control of playback of the source video information.
- instant access to any location within the video information.
- the ability to compare multiple video and audio tracks simultaneously.
- the ability to activate any track instantly.
- zoom controls for in-depth analysis of video information.
- integrated note-taking that is accurate to the level of one video frame.
- user-definable categories for the note-taking process.
- user-definable time resolution for the note-taking process.
- assistance in speeding up the note-taking process.

These points will be discussed in the following sections as we begin looking at the features of *Video Transcriptor* in detail.

Section 2

Installing *Video Transcriptor*

System Requirements and Installation

System Requirements

To use the *Video Transcriptor* program, you need:

- A Macintosh® II (or any later model) computer that is QuickTime-compatible, with a minimum of 4 megabytes (MB) of random-access memory (RAM).
- A video display card that is capable of displaying 4 bits (16 colours) or more.
- A hard disk drive with at least 40 MB of storage.
- Apple® System 6.0.7 or later.
- 32-Bit QuickDraw™, version 1.2 or later.
- QuickTime™, version 1.6 or later.
- Access to a video digitising board that produces movies in the QuickTime format. This is required to capture the time-based data into a form readable by the program.

Performance improves with more memory, a faster CPU, and faster and larger hard disk drives.

Installing Video Transcriptor

The installation process for *Video Transcriptor* is routine.

- Make a folder on your hard disk drive and label it **Video Transcriptor.**
- Copy the application *Video Transcriptor* from the floppy disk into this folder.
- Copy the QuickTime System Extension to your System Folder if it is not already there.
- Copy the two movies "Sample Movie" and "Sample Movie Two" from the floppy disk into this folder.
- Restart your computer for the QuickTime System Extension to take effect.
- Open the *Video Transcriptor* folder and double-click the *Video Transcriptor* program icon.

The Master Controller window is opened automatically; . The process of note-taking may now start. For a quick tutorial on how to use *Video Transcriptor* refer to the "Quick Start" section at the beginning of this manual.

Changing the application memory size

The *Video Transcriptor* program is set to an application memory size of 1500 KB. You can improve the program's performance by increasing the application memory size. To change the application's memory size:

- Make sure that *Video Transcriptor* is not currently running.
- Open the *Video Transcriptor* folder on your hard disk drive, and click the *Video Transcriptor* icon.
- Choose *Get Info* from the File Menu. The Info box will appear.

- In the Application Memory Size box in the lower right corner, enter the maximum amount of memory that you can allocate to the *Video Transcriptor* program.

Note: This must be less than the free memory available.

System 7 and 32-bit addressing

Video Transcriptor is a "32-bit clean" application, which means that it is designed to use the 32-bit addressing capability of System 7. This allows it to access more than 8 MB of RAM, so larger files can be stored entirely in memory and less hard disk drive access is required.

Some of the earlier Macintosh computers do not support 32-bit addressing as their ROM (read-only memory) chips are not 32-bit clean. These machines are those prior to the Macintosh fx and ci models. On Macintosh models prior to the fx and ci, a third-party utility such as *Mode-32* can be used to run in 32-bit mode.

Section 3

Tutorial

Introduction

This Section shows you how to use the facilities available in *Video Transcriptor*. It is designed to be a self guiding tutorial of the features available. Later sections will discuss more advanced features that *Video Transcriptor* provides.

Topics covered in this section:

- Starting *Video Transcriptor*
- Opening a Movie
- Playback Controls
- Volume Controls
- Controlling Movie Windows
- Other Features of the Master Controller
- Video and Audio Windows
- Transcription Process
- Menus

This tutorial is designed to give you a full understanding of the features available in *Video Transcriptor* while analysing the sample QuickTime movie "Sample Movie Two" which has been supplied. For a brief overview of how to use *Video Transcriptor,* see the section "Quick Start" at the beginning of this manual.

Before analysing video and audio data on a computer the information must be in a format that the computer can read. This format for the video and audio information is a "**QuickTime Movie**". *Video Transcriptor* does not support the creation of these files. The digitisation process of the analog video and audio signals into the digital format that the computer requires can be achieved with a third-party video digitising card. These digitising cards usually comes with the required software to record the video clip into the QuickTime movie[1] format or you can use many free or shareware programs that are available. Once the video clip is in the QuickTime movie format, *Video Transcriptor* can be used.

Starting Video Transcriptor

Open the *Video Transcriptor* folder on the hard disk drive and double-click the *Video Transcriptor* program icon.

Upon successfully starting the application you should see a window in the top left hand corner of the screen, as shown in figure 3.1. This is the "Master Controller" window which is used to control the playback of the video and audio information in the QuickTime movie.

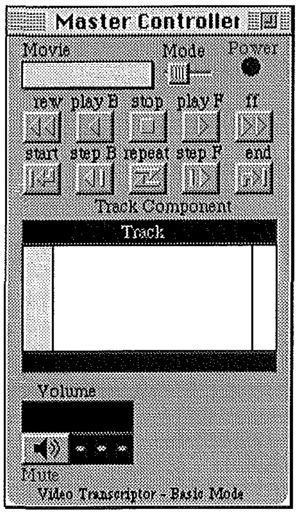


Figure 3.1. The Master Controller Window.

[1] A **movie** is the QuickTime file format that contains references to multiple video and audio tracks.

In order to start the analysis process, a movie must first be loaded to work with. In this tutorial section the "Sample Movie Two" QuickTime movie has already been created.

Opening The Source Movie

To open the movie, select the *Open Movie* command from the File menu. Figure 3.2 shows the file selector box which appears.

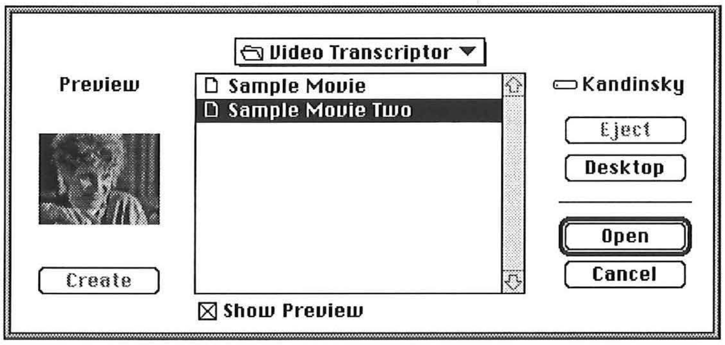


Figure 3.2. The Standard Macintosh File Selector Box.

Select the "Sample Movie Two" movie file and press the *Open* button. This will load the movie file ready to be analysed. If the movie has been loaded correctly, two more windows should appear above the Master Controller window; one containing video images, figure 3.3, the other containing information about the soundtrack of the movie, figure 3.4.

Figure 3.3. The Video Window.

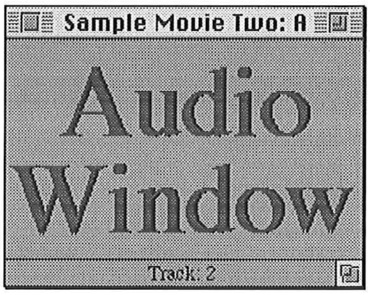


Figure 3.4. The Audio Window.

The Master Controller window has also changed. It now contains information relating to the movie just loaded, figure 3.5.

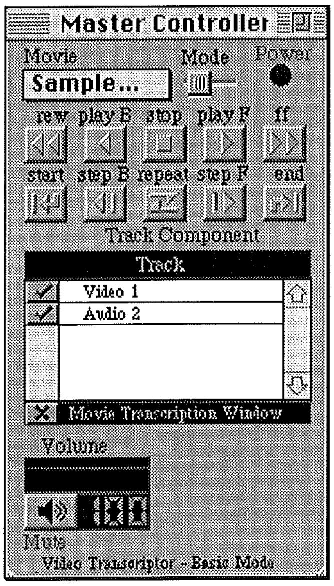


Figure 3.5. Master Controller After Loading in *Sample Movie Two*.

It contains the name of the movie file loaded in a pop-up list in the top left corner, information about the video and audio track loaded in the centre, and volume information at the bottom of the window. More details about the functionality of these elements will discussed later in this section.

Playback Controls

Overall control of the movie data is handled by several controls on the Master Controller. Figure 3.6 shows the controls that are available to assist in the manipulation of the movie data. These controls are located at the top of the Master Controller window.

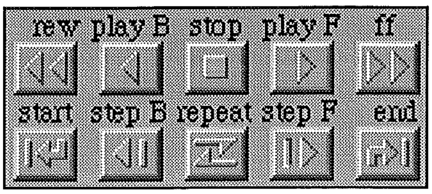
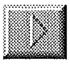
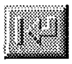


Figure 3.6. Movie Controls.

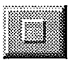
The top row of buttons, together with the "step B" and "step F" buttons, represent the standard features that can be found on a video-tape player's remote control. The remaining three buttons, "start", "end", and "repeat", offer additional features that the video-tape player does not offer.



Press the "play F" button and watch the movie being played back. You will notice that when the movie reaches the end it stops automatically.



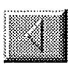
Press the "start" button to instantly return to the start of the movie. On a video-tape player this action would take several minutes while the tape is rewound to the beginning. Since the video data is stored on the computer in a digital format, any location in the movie is accessible instantly.



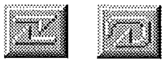
After pressing the "start" button, press the "play F" button again to start playing the movie. Press the "stop" button while the movie is playing. This halts the movie where it is. Unlike a video-tape player, the picture does not disappear but remains as it last appeared before the "stop" button was pressed. It acts like the pause button on a video-tape player.

Try using these three buttons to get the feel of playing the movie back and stopping it as required. When you are familiar with these controls press the "start" button to return to the beginning of the movie. Try pressing the "return" key. This has the same effect as pressing the "play F" button. To stop the movie try pressing the "space bar" instead of the "stop" button. These keyboard shortcuts are available to assist the user in controlling the playback of the movie without always having to rely on using the mouse. Once you feel comfortable with these buttons and their corresponding keyboard shortcuts we can move on to the remaining buttons.

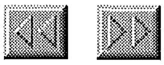
Press the "end" button to move to the end of the movie. This is equivalent to the "start" button discussed before.



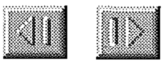
Press the "play B" button to play the movie in reverse. This will also play the audio track in reverse. When the movie reaches the start it will stop.



Try pressing the "repeat" button. The button changes shape from the two directional arrows into a single circular arrow. Press the "play F" button and watch what happens when the movie gets to the end this time. You will notice that it starts over again from the beginning, looping forever until it is stopped by pressing the "stop" button which will stop it immediately, or by pressing the "repeat" button which will stop it when it next reaches the start or end of the movie, depending on the direction that it is being played back in.



Other controls similar to those found on a video player's remote control, are the "ff" and "rew" buttons. These will fast forward and rewind the movie at a high speed in either direction. Halt the movie with the "stop" button somewhere in the middle. Then try using the "ff" and the "rew" buttons to see how they work. Review all of the functions of the buttons that you have learnt so that you know exactly what they do.



The last two buttons are very useful in analysing sequences of video data. These buttons will step the movie through frame by frame. Use the "start" button to take the movie to the beginning and the press the "play F" button to start playing the movie. Use the "stop" button to stop the movie somewhere in the middle. Try using the "step F" button a few times to step through the video frames. Use the "step B" button to step backwards a single frame. Alternatively you can use the two arrow keys, left and right, to step through frame by frame.

With these controls you will be able to reach any location within a movie in order to analyse the movie data in detail during the note-taking sessions.

Controlling the Volume

The overall volume for the movie is set via a control found on the Master Controller window. This control, found at the bottom of the Master Controller window, is illustrated in figure 3.7.

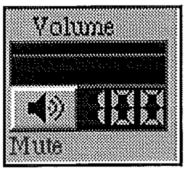
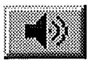


Figure 3.7. Volume Controller.

There are two controls available; the mute button and the volume level slider.



The mute button terminates all audio output until the mute button is depressed again. When the audio is muted the button changes to the following.

The volume is controlled by a horizontal sliding bar. Try pressing the mouse button anywhere in the volume area. This will change the volume setting relative to the position of the mouse. By moving the mouse over the right-hand edge of the blue volume slider bar, the cursor changes shape into a red two directional arrow. Pressing the mouse button down and drag the mouse horizontally. This changes the volume up and down accordingly. During this activity, the volume level is updated below the bar in green numbers ranging from 0, for no volume, to 100, for full volume. Figure 3.8 illustrates changing the volume level by dragging the volume level slider.

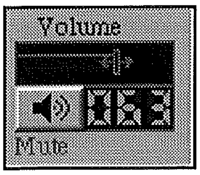


Figure 3.8. Changing the Volume Level.

When the volume is muted the volume level is indicated by "---" and the volume bar changes colour to grey, indicating the volume control cannot be used until the mute button is depressed again, figure 3.9.

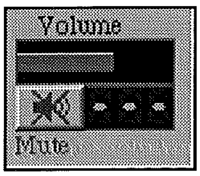


Figure 3.9. Volume Control When the Movie is Muted.

Controlling Movie Windows

Each movie that is loaded may contain many windows. These windows may contain video information, audio information or notes about the movie. The Master Controller window contains methods of controlling these windows from a single location. This control is located in the middle of the Master Controller window. Figure 3.10 shows the Track Component controls responsible for controlling the various windows.

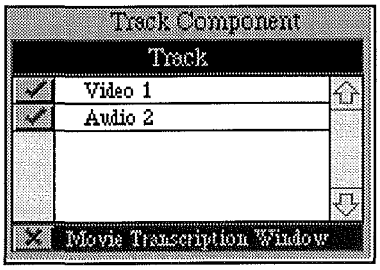


Figure 3.10. Track Component.

Each row in the list represents a track that is in the movie. In the "Sample Movie Two" movie there are two tracks; a video track labelled "Video 1" and an audio track labelled "Audio 2". At the bottom there is a button for opening the movie's transcription window.

Each row contains a button for activating the track and displays the type and the number of the track found in the movie. The button contains a tick or a cross which indicates if the track is currently open and activated. The type of the track is either a video or an audio track. The numbering has been incorporated to assist the user in keeping track of items related to each track on the screen.

Close the video window for the movie "Sample Movie Two" by pressing the button beside the track labelled "Video 1". The tick on the button changes into a cross indicating that the track is no longer open and active. The ability to turn individual tracks on and off can be very useful when analysing movies that contains many tracks. To bring the track back press on the button again. The button changes back to a tick.

Try double-clicking on the *Track* entry labelled "Video 1" beside the button. If the window is already open but hidden behind other windows then by double-clicking the window will be brought to the front of all the other windows that are open. If the window is not already open then it will be opened.

The button at the bottom of the Track Component labelled "Movie Transcription Window" will open the transcription window that is used for taking notes about events for the movie. Try pressing the button to open the transcription window for the movie. This will be discussed later in this section under "Transcription Process". Press the button again to close the window while we examine some of the other features of *Video Transcriptor.*

Other Features of the Master Controller

We will look at some of the remaining features that can be found on the Master Controller in *Video Transcriptor.*

Power Indicator

The "Power" light, located at the top of the Master Controller window, is representative of a light on a video player's remote control. When the user presses a button on the remote control, such as the play button, the light is turned on until the button is released. The purpose of the power light is to indicate a response to a users action. Figure 3.11 shows the light in its off state and figure 3.12 shows it when it is activated.

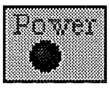


Figure 3.11. Power Light - Off.

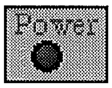
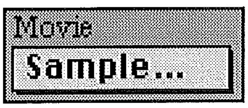


Figure 3.12. Power Light - On.

Movie Indicator



The movie indicator is a pop-up list of any movies opened by the user. The user can load any number of movies, each containing multiple tracks. To switch between the different movies the user can either click on an inactive window from another movie or select another movie from the pop-up list shown in figure 3.13.

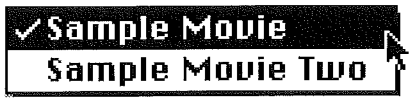


Figure 3.13. Pop-up List of Active Movies.

By changing movies, all the information on the Master Controller is updated to reflect this change of movie data.

Mode Information

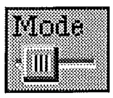
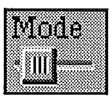
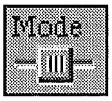


Figure 3.14. Mode of the Movie.

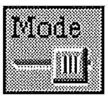
The Master Controller has three modes of operation; basic, standard and expert. The current mode is available at the bottom of the Master Controller. The mode can be changed in the *Options* which is available in the Edit Menu or by clicking in the slider control shown. The different modes affect the controls available on the Master Controller. You will notice that the movie's current mode is on the leftmost setting. This implies that *Video Transcriptor* is set to the "BASIC" mode. As you become more familiar with the features available in this mode then you may wish to move up to the next mode. Below is a brief description of the facilities available in each of the three modes.



The *basic* mode contains the bare essentials for movie playback and manipulation. Only the standard movie controls as found on a video remote control unit are available. The control for variable speed playback is not available along with positioning controls. The zoom controller is also unavailable by default. However it can be selected in the *Options* item from the Edit Menu. Only one transcription window is available for taking notes about the movie as a whole. Individual transcription for each track in the movie is not available until the STANDARD mode. The *basic* level is intended as a bare minimum set of playback controls for the movie. It removes all enhanced modes until the user is more confident with the initial ideas behind the manipulation of the movie. The user may then change the level of controls available to a higher level.



The *standard* mode contains all the controls except for the variable speed controller. This is intended for users who are more confident in controlling the playback of the movie. The *standard* mode introduces more enhanced facilities for controlling the movie data. These include the introduction of the zoom controller, positioning controller and more audio controls on the Master Controller. The STANDARD mode also introduces multiple transcription windows for each track in a movie.



The *expert* mode contains the full set of controls that *Video Transcriptor* supplies. This is intended for users who are fully comfortable with all of the features available for controlling the playback of the movie.

Iconise

There are times when the Master Controller takes up too much room on the desktop. The zoom button in the right-hand corner of the title bar of the Master Controller acts as a button to "Iconise" the Master Controller on the desktop. Try pressing this zoom box to see what happens. Figure 3.15 shows the result of what happens to the Master Controller.

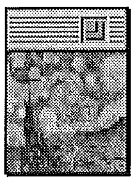


Figure 3.15. Master Controller Iconised.

Click in the zoom box again, or click anywhere in the contents of the iconised window. The Master Controller will be restored back to its original position.

Option Key

Try holding the option key down while clicking in the zoom box this time. The iconised form will be positioned at the top left hand corner of the screen. Try holding the option key down while restoring the iconised controller back to its original size. The Master Controller will be positioned at the top left hand corner of the screen.

Video and Audio Windows

The video and audio windows as shown below in figures 3.16 and 3.17 contain the visual information about a movie.

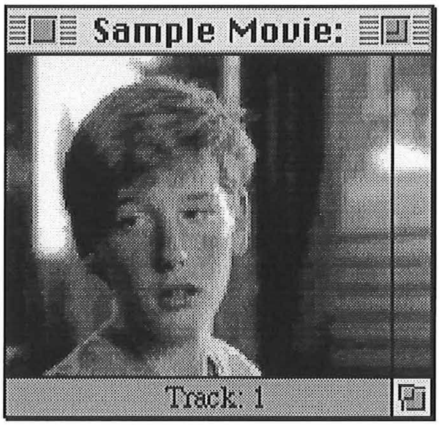


Figure 3.16. The Video Window.

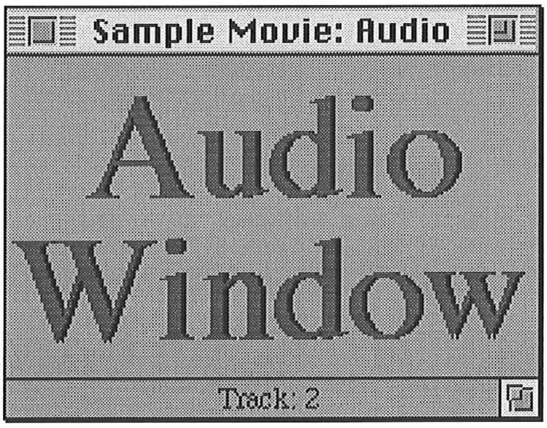


Figure 3.17. The Audio Window.

The Video window contains the movie's picture. Try resizing the window to enlarge the picture. This is useful to get a better view of the whole movie. All video windows in the movie are also resized to the same size. By enlarging the picture the playback results may be affected however. The larger it is resized the slower the playback speed may become.

If the window is not needed it can be closed. It can be retrieved as discussed earlier in the section "Controlling Movie Windows".

The audio window is just representative of the audio track being currently active. Enhanced capabilities are available under the *Expert* mode in *Video Transcriptor,* see Section 5 - *Expert Mode Features.*

Transcription Process

By this stage you should be familiar with controlling the playback of the movie. The next stage in this Tutorial is the note-taking process. To begin the note-taking process select *New Transcript* from the Transcript Menu, figure 3.18.

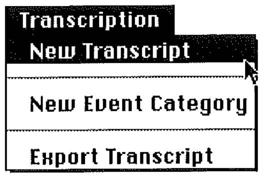


Figure 3.18. Selecting *New Transcript* from the Transcript Menu.

This will bring up the movie transcription window for taking notes, figure 3.19.

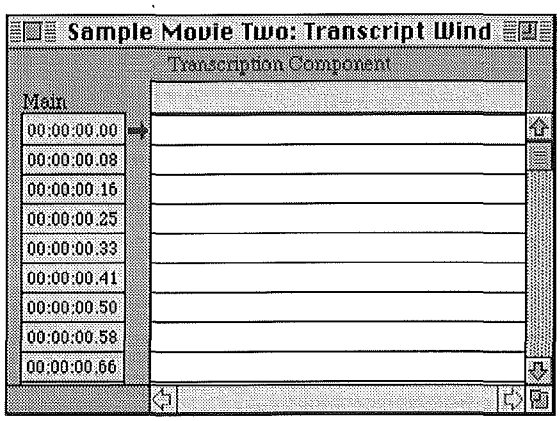


Figure 3.19. The Transcription Window.

There are three main elements in the transcription window; the time resolution of the movie data, user definable event categories, and the actual events entered by the user. Figure 3.20 shows where the main elements of the transcription window are located.

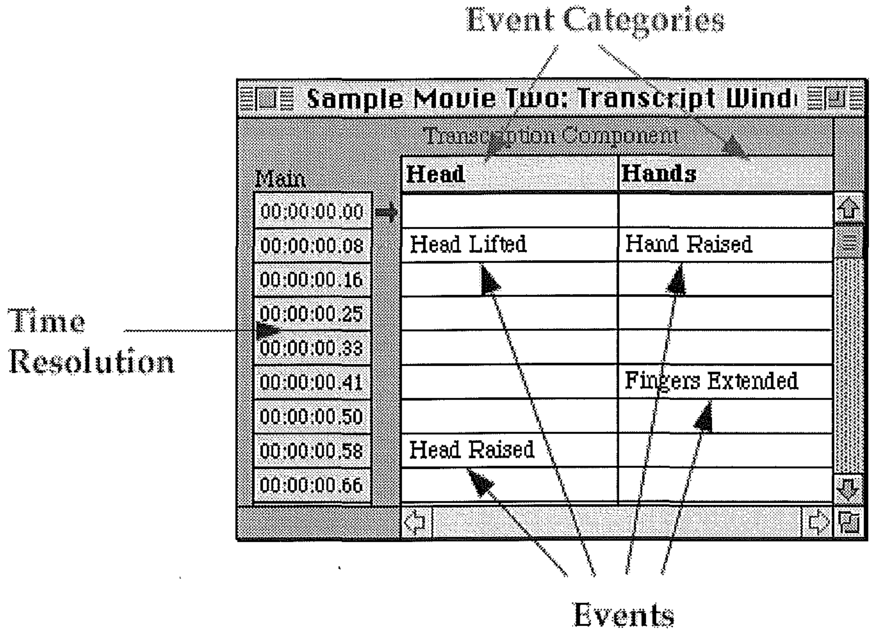
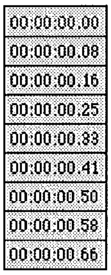


Figure 3.20. Elements of the Transcription Window.



The box to the left hand side of the window shows the time resolution of the movie. Each interval represents the time of an individual video frame in the movie. In this example we can see that each interval is 0.08 seconds in length. The format of the time is *hours:minutes:seconds.hundredths of a second*. The current time of the movie is highlighted in red and with an arrow pointing to any events that have occurred at this time. This lets you see at a glance any notes that have occurred at each point in the movie.

The box at the top of the window contains the event categories that are definable by the user. The user definable **event categories** define what the notes are about. In figure 3.20 two groups have been defined; one for notes about the subject's "**Head**", and the other for notes about the movement of "**Hands**". These categories define the subject of the notes in each column. For example everything below the **Head** category lists details about the subject's head movement in the movie. The same applies for the **hand** category. Any notes below it are related to arm movement by the subject.

Head Lifted	Hand Raised
	Fingers Extended
Head Raised	

The majority of the window is taken up by the table of events . The table of **events**, as shown above in the transcription window, contains the notes about events that have occurred. Each event occurs within a time interval in the movie and is contained within a defined event category.

For example, in figure 3.21 below, we can see the event "Hand Raised" occurred at time 00:00:00.08. This event indicates that the subject moved their hand upwards in some fashion. We can determine from the event category what the event is about. We will now construct the following notes about the movie "Sample Movie Two" as we can see in figure 3.21.

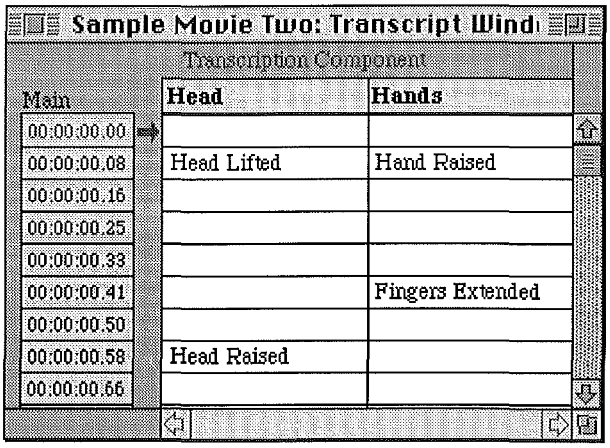


Figure 3.21. Transcription Window Containing Events.

After playing the movie through a few times using the controls on the Master Controller window, we can get some idea of the events which have occurred and the type of notes to be taken. From this we can examine the process of creating event categories and the procedures for maintaining these groups.

Event Categories

- **New Event Categories**

First of all, we will define some **event categories** for the movie. These event categories are used to organise our notes into similar event groups.

We will define the two event categories; "Head" and "Hands". To create the first of these event categories hold down the mouse button while pointing to area where the event categories are displayed. This brings up a pop-up menu, figure 3.22.

Figure 3.22. Pop-up Menu For New Event Categories.

You can only select ***New Event Category*** from the pop-up list at this point. This will bring up a window to enter the category name, figure 3.23.

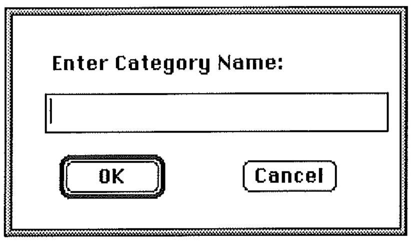


Figure 3.23. Dialog Box For Entering Event Categories.

Enter "Head" into this window and then press the *OK* button.

We will create the second event category in a different way this time. This time select *New Event Category* from the Transcript Menu, figure 3.24.

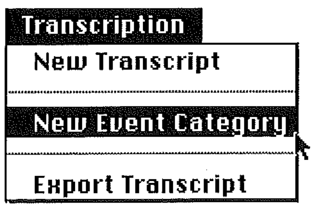


Figure 3.24. Selecting *New E vent Category* from the Transcript Menu.

This will open the same window as we saw in figure 3.23. Enter "Hands" this time as the event category name. We should now have two categories defined; Head and Hands.

Two categories should now exist in the transcription window for notes to be entered. The transcription window now look like figure 3.25.

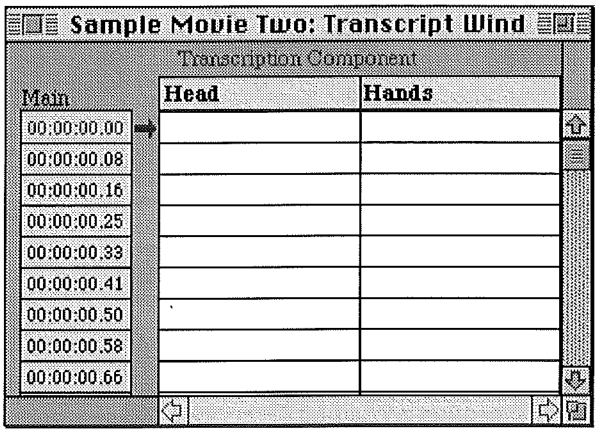


Figure 3.25. Transcription Window With Two Event Categories Defined.

Try adding some more event categories if you wish to take some notes on different aspects of human behaviour in the "Sample Movie Two" movie file.

- ### Modifying Event Categories

Try double-clicking on one of the event categories. This brings up a window, figure 3.26, with the contents of the particular event category.

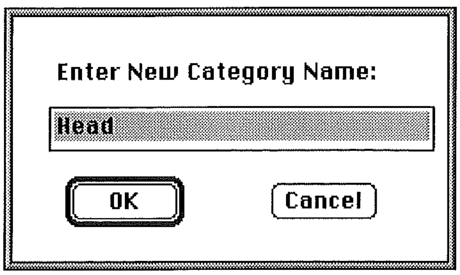


Figure 3.26. Modifying an Event Category.

From here we can change the category name or even fix a spelling mistake if any had been made.

Try holding the mouse button down while pointing to a different event category. This will bring up a pop-up list , figure 3.27.

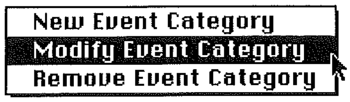


Figure 3.27. Event Category Pop-up Menu, Modify Event Category.

Select ***Modify Event Category.*** This once again brings up the window in figure 3.26 to change the contents of the event category.

- ## Removing Event Categories

To remove an event category and all of the events associated with the particular event category, hold down the mouse button on the particular event category. This will bring up the same pop-up menu that we saw for modifying event categories, figure 3.28.

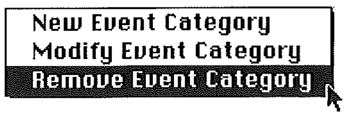


Figure 3.28. Event Category Pop-up Menu, Remove Event Category.

By selecting ***Remove Event Category*** a second window appears asking the user to confirm that they wish to remove the event category and all of the events associated with it, 3.29.

Figure 3.29. Confirm Deleting and Event Category.

Note: This is a very powerful feature that should be used wisely and with considerable thought. Once you have confirmed that you wish to remove the event category and all of the events associated with it, you will be unable to recover this data.

Events

Once you have created the event categories that organise your data, you can then begin to create entries on the individual events that occur during the movie.

Each entry in the events table that can potentially contain information about an event that may have occurred during the current time interval. During one time interval many potential events may occur. For example an event concerning the subject's head may occur together with an event about the subject's hand. Under this situation the column "Head" and "Hands" will both contain an entry that occurred at the same time.

Under normal conditions the individual events would be scattered throughout the events table. However if the analysis was of a different nature, a high speed chemical reaction for example, then potentially all of the cells in the events table might be filled with entries.

We will attempt to recreate the events that are visible in figure 3.24. This will give you all the information needed to continue examining the rest of the movie.

- **Creating a New Event**

Our first event that we are going to produce is located at time 00:00:00.08 in the movie. Use the video-like buttons to relocate the movie to this new location. The "step F" button is the most appropriate button in this situation. If you step forward a few more frames and watch the video window while doing this, you will notice that the subject's head was only slightly raised. Take the movie back to the location at 00:00:00.08 by using the "step B" button. In figure 3.30, the same as figure 3.24, you can see the entry in the "Head" column at time 00:00:00.08.

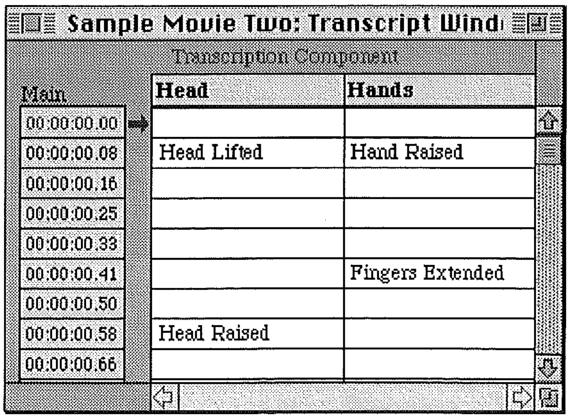


Figure 3.30. Transcription Window Containing Events.

This entry is titled "Head Lifted" which corresponds with the first frame when the subject's head was beginning to move upwards.

To enter details about this event in our transcription window, double-click the mouse button in the cell located beside the time 00:00:00.08 and in the event category "Head". This brings up a window for the user to enter details about the event, figure 3.31.

Figure 3.31. Window for Entering New Event Information.

At the top of the window the current movie time is shown. There are two fields for the user to fill in. A short description about the event must be entered. In our example, in figure 3.31, we had entered "Head Lifted". Enter this into the short description field and either press the *return* key, the *tab* key or click in the large text entry box below the short description field with the mouse. This field contains any additional notes that are appropriate for this event. This field is optional so you do not have to enter anything here. We cannot see what had been entered in this field from figure 3.30 without getting more information about the individual event. The purpose of the short description field is to convey as much information about the event possible. For this event enter "Head raised in response to moving hand" into the longer description field.

When you are finished entering the details about the event then you can press either the *OK* button to record the event details, or press the *Cancel* button if you wish to abandon the new event.

If you select *OK* without entering a short description then the alert window in figure 3.32 is displayed.

Figure 3.32. Alert Window When No short Description is Entered.

We shall now enter the next event that is available. From figure 3.30 we see that this is event is also located at time 00:00:00.08. You may wish to examine the video frames leading up to and beyond this event so that you can see what the event is all about. When you are ready to proceed, click and hold down the mouse button in the cell at time 00:00:00.08 in the event category "Hands". After a short time this will bring up a pop-up menu, figure 3.33.

Figure 3.33. Event Pop-up List With No Previous Events in the Event Category.

You must select *New Event* from the pop-up list at this point. Once again this will bring up the window as seen in figure 3.31. Enter "Hand Raised" into the short description field and "Hand raised to point at object" in the long description field. When you have finished, press the *OK* button.

Move to the position that the third event is located in the movie, at time 00:00:00.41. Once again use the movie controller buttons to find out what the event is about. When you are ready to continue click and hold down the mouse button in the cell at time 00:00:00.41 in the "Hands" event category. After a short time a pop-up menu will appear as in figure 3.34.

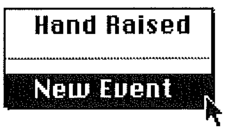


Figure 3.34. Event Pop-up Menu.

This pop-up menu has two items in it; the first item is labelled "Hand Raised", the second labelled "New Event". This pop-up menu is divided by a horizontal line to separate the two sections. Any items in the top section is a unique list of all the short descriptions that have been entered previously within the event category.

We will look at the top section of the pop-up menu next.. Select "New Event" from this pop-up list.

The resulting window that will appear can be seen in figure 3.35.

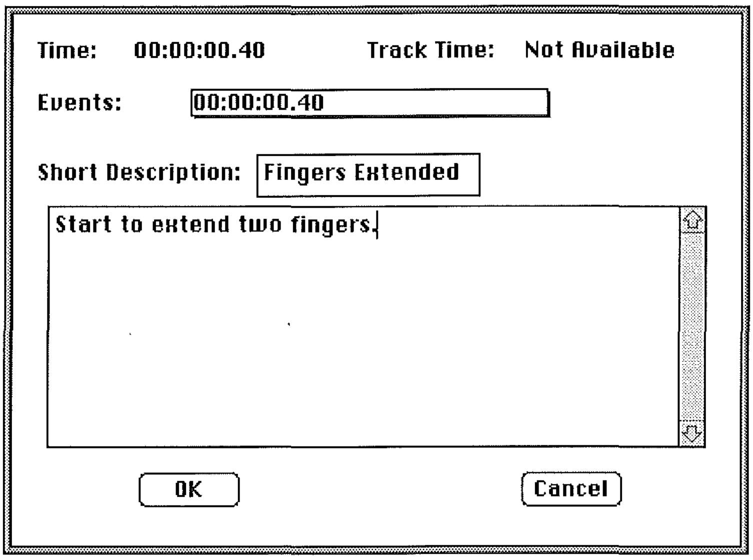


Figure 3.35. Window For Entering New Event Information.

This is the same window as we saw before in figure 3.31. Enter "Fingers Extended" into the short description field and enter "Start to extend two fingers" into the longer description field. When you are finished press the *OK* button to record the event details.

We shall now move onto entering the last of our four events. Move to the position that the last event is located at, 00:00:00.58, refer to figure 3.30. Click and hold the mouse button down while pointing to the cell in the "Hands" event category. A pop-up menu appears that is similar to the one in figure 3.34. The one that appears, figure 3.36, contains "Head Lifted" in the top section where "Hand Raised" was visible in figure 3.34. "Head Lifted" is visible since it is the only other event already defined within this event category. "Hand Raised" is in the "Hands" event category.

Figure 3.36. Event Pop-up Menu.

It does not list any items in the other event categories like the "Hands" event category for example.

This time select "Head Lifted" from the pop-up menu. You will notice that it doesn't match the entry in figure 3.30 but this is to illustrate the point of modifying an event which is discussed in the next section.

Instead of a window, figure 3.31, appearing for details about the event, the event is automatically inserted into the events table. You will see that at time 00:00:00.58 the new event labelled "Head Lifted" has been inserted. Later we will see that only the short description field has been filled in about the event. The longer description field is empty.

As more new events are created within an event category, the pop-up menu will grow in size. The top section of the pop-up menu will contain a single entry for each unique short description entered about an event.

When taking notes, the short descriptions for the individual notes will normally be duplicated. For example in our "Head" event category we may end up with four short descriptions relating to head movement; move left, move up, move right, and move down. If these events occur frequently then it would be inefficient to have to enter this text in all the time into the window in figure 3.31. It would be more efficient for the computer to build up a list of previous events that have occurred, allow the user to optionally select from this list of events, and have the event details automatically recorded. If the user wants to create a new event type then selecting "New Event" or double-clicking in the event cell would be appropriate.

- **Modifying an Event**

There are two ways to modify an event that has been entered into the transcription window; by double-clicking in the event cell or by the pop-up menu.

Double-Clicking Method

Try double-clicking in the cell that the last event was entered into at time 00:00:00.58. Previously the wrong description was entered about the event. The same window used to enter the information about the event appears, figure 3.37.

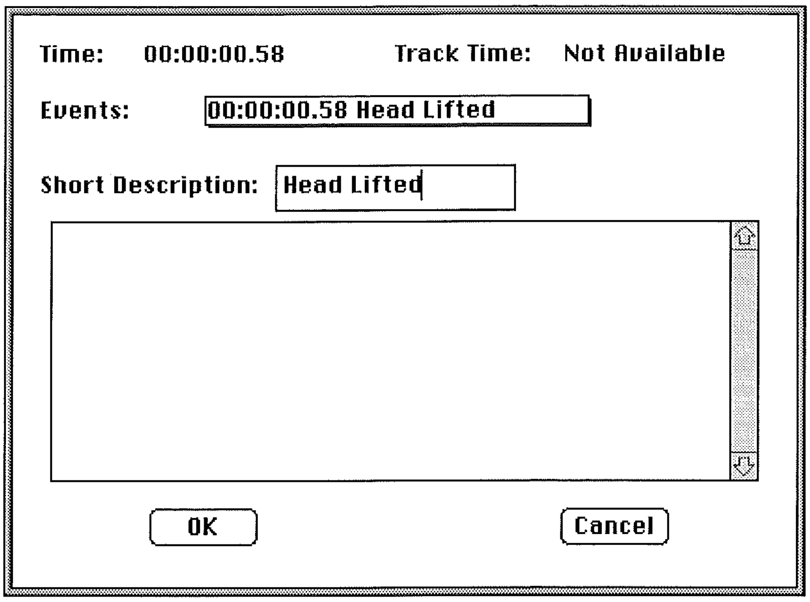


Figure 3.37. Window For Modifying an Event's Contents.

This window contains the information that had been entered when the event was created. Change the short description field to read "Head Raised" and press the *OK* button when finished.

Pop-up Menu Method

The second method available is the pop-up menu. Figure 3.38 shows the pop-up menu when selecting the event which was just modified.

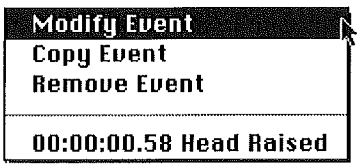


Figure 3.38. Pop-up Menu For Modifying an Event.

You will notice that the format of the pop-up menu has changed. Instead of having the top section of the pop-up containing a unique list of event titles that have occurred, it now contains a list of actions that can occur to the event. We are not trying to add a new event but are wanting to perform an action on it.

There are three actions available; modify an event, copy an event, and remove and event. These last two items will be discussed in the following sections.

Selecting the "Modify Event" item will bring up the window to modify the event details, figure 3.39.

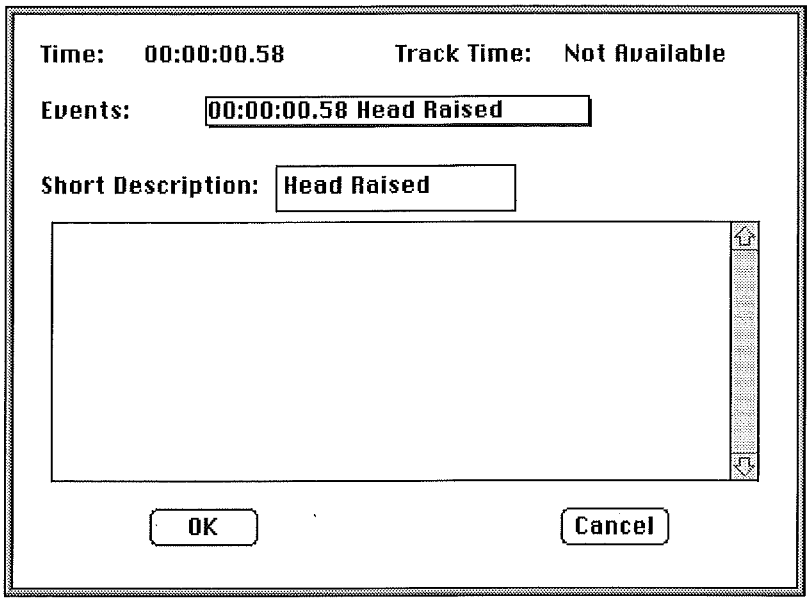


Figure 3.39. Modifying an Event's Contents.

Last time we modified this event we just changed the short description so that it was the same as figure 3.30. This time we will add the long description about the event. By automatically inserting the event by using the items in the top portion of the pop-up menu, only the short description is added. The long description still has to be added manually if it is required. For this event enter "Subject begins talking" into the long description field.

- ## Removing or Moving an Event

From the pop-up menu in figure 3.38 we can see the menu item for removing an event. We shall remove the last event that we created and later recover it using the "Paste Event" menu item. Click in the "Head" event category and bring up the pop-up menu available at time "00:00:00.58", figure 3.40.

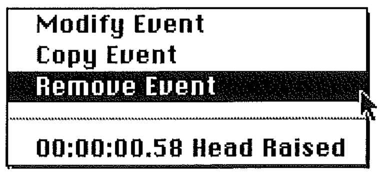


Figure 3.40. Pop-up Menu for the Event at Time "00:00:00.58" From the "Head" Event Category.

Select "Remove Event" from the pop-up menu. This will remove the event from the events table and copy it into a special clipboard buffer. This acts just the same as the "Cut" menu item in a word processor. The actual contents can be recovered again by selecting "Paste".

This command can be used in two ways; as a step in moving the event to a new time location or even into a different event category, the second function totally removes the event from the transcription window.

The purpose of the special clipboard buffer that the event details are copied into is to temporarily store the contents of the event that has just been removed. This allows the user to recover the event that was just "removed" in case it wasn't meant to be removed. Since it is still in the special clipboard buffer it can easily be recovered using the "Paste Event" command discussed in the following section. The user can also paste the event stored in the special clipboard buffer to any new location in the events table. The next time an event is "Removed" or "Copied", the contents that were previously in the special clipboard buffer are lost forever being replaced by the new event.

- **Pasting an Event**

Since we didn't really want to remove the last event we will recover the contents back to its original location.

Click and hold the mouse button down in the "Head" event category at time "00:00:00.58" to bring up the pop-up menu, figure 3.41.

Figure 3.41. Paste Event pop-up Menu.

Select "Paste Event" from the pop-up menu. This will copy the contents of the special clipboard buffer into the selected cell. We now have the contents of the transcription window looking as the original in figure 3.42.

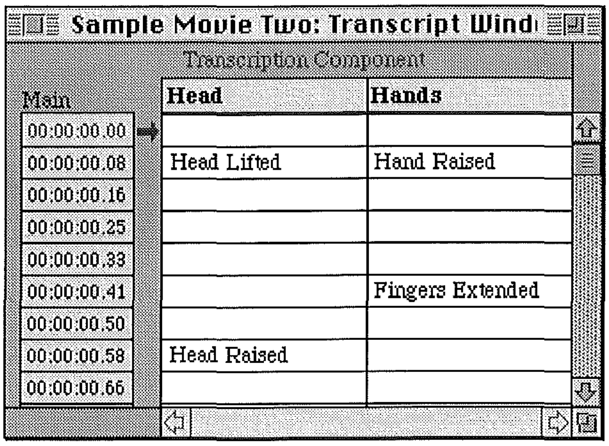


Figure 3.42. Transcription Window With Completed Event Entries.

The contents of the removed event are still available in the special clipboard buffer. This may be repeatedly pasted to many locations if needed.

- **Copying an Event**

We have already seen a form of this command before in the section "New Event". In this situation it is possible create an event from those available in the pop-up list. This method would create a new event which contains the short description that had been defined for another event within the same event category. This method speeds up the transcription process for repetitive tasks, figure 3.43.

Figure 3.43. Inserting a New Event Automatically.

The "Copy Event" command in comparison makes an exact duplicate of all the available information which can the be pasted to many new locations. Click in the "Hands" event category at the time "00:00:00:41" and hold down the mouse button until the pop-up menu appears. Figure 3.44 shows the pop-up menu.

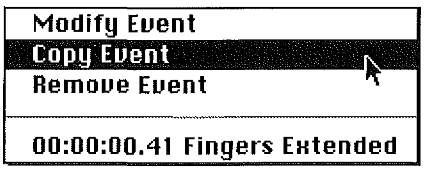


Figure 3.44. Pop-up Menu Used to Copy an Event.

Select "Copy Event" from the pop-up menu. This makes a copy of the event into the special clipboard buffer. The contents of this buffer can then be pasted to any location in the events table.

Try moving to any cell that doesn't already contain an event, time "00:00:00.66" in the "Hands" event category for example. Select "Paste Event" from the pop-up menu, figure 3.45.

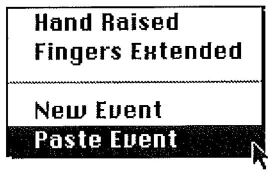


Figure 3.45. Pasting an Event into an Empty Cell.

The contents of the special clipboard buffer will be copied into this new cell. Remove this event by selecting "Remove Event" from the pop-up menu that now appears when you select the event that you just inserted, figure 3.45.

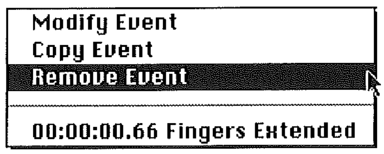


Figure 3.46. Removing an Event.

You may wish to try analysing the rest of the "Sample Movie Two" movie file to become more familiar with the movie commands that are available.

Pop-up Menu States

A summary of the pop-up menus for the transcription window is shown below in figure 3.47 for event category pop-up menus and 3.48 for event pop-up menus.

The layout of the pop-up menus for event categories is below in figure 3.47.

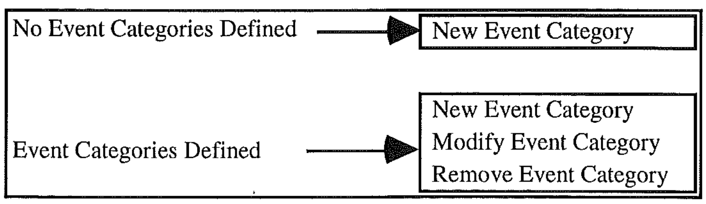


Figure 3.47. Event Category Pop-up Menus.

When no event categories have already been defined then the pop-up will only contain the option *New Event Category* that we saw previously in figure 3.19. When event categories have been defined then you can either; create, modify (figure 3.27), or remove an event category (figure 3.28).

The layout of the pop-up menus for events is below in figure 3.48.

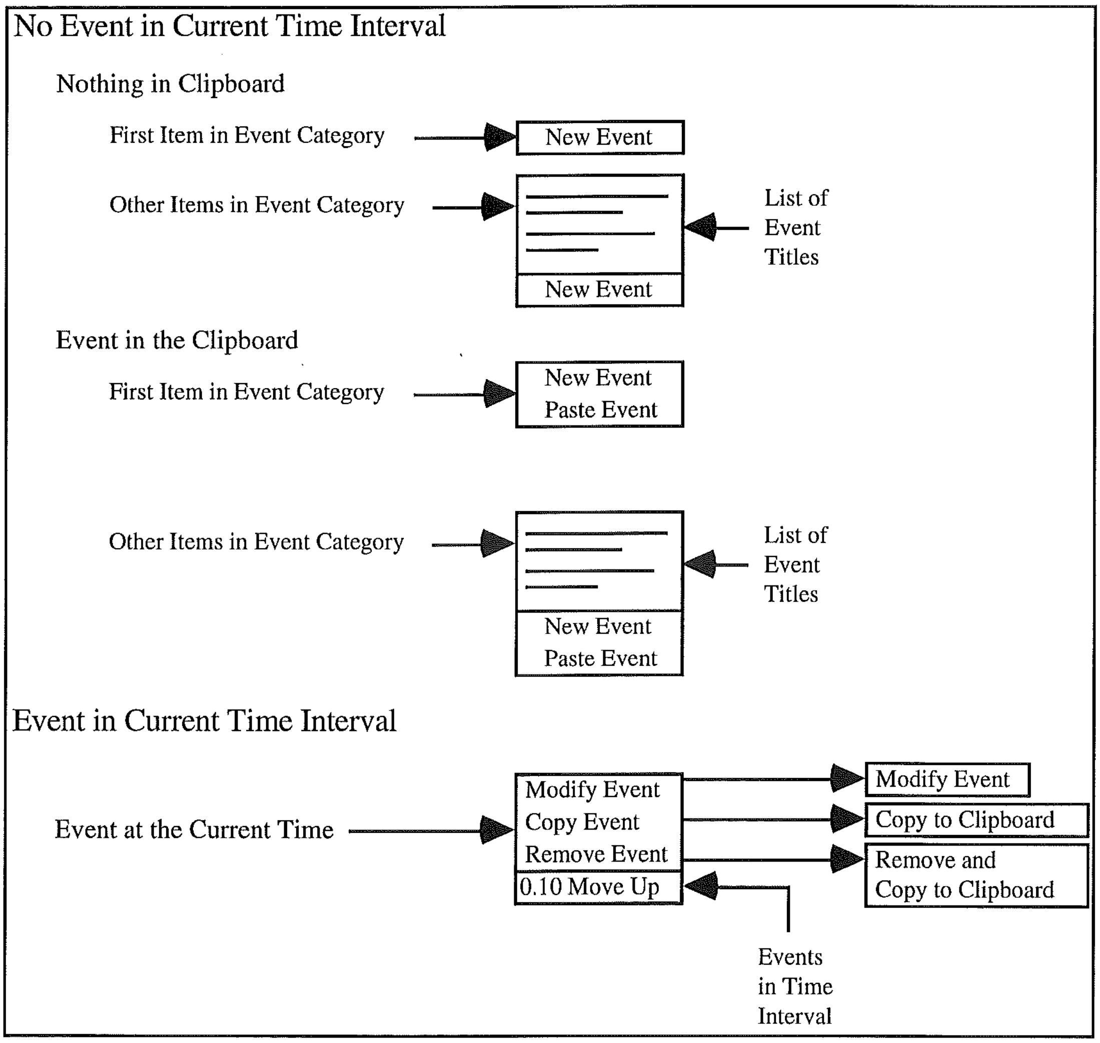


Figure 3.48. Event Pop-up Menus.

There are two situations that can occur, either there is an event at the current time or it is empty. If it is empty then we can once again spit this decision into two once again; either there is something in the special clipboard buffer or it is empty. Figure 3.48 breaks up the menu choices into these different categories to make it easy to follow.

The right hand side of the description illustrates the format that the pop-up menu should take when it is invoked.

Saving Results

Now that you have finished the note-taking you will want to save the results so that you can retrieve them next time you use *Video Transcriptor*. Select "Save" from the File Menu, figure 3.49.

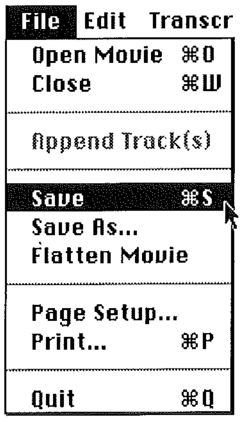


Figure 3.49. Selecting Save From the File Menu.

This will store any notes taken with the movie file. When you quit from *Video Transcriptor* and come back to use it at a later date to examine the current movie file, all of the notes taken will be preserved.

Exporting Results

Now that you have finished taking notes about the movie and saved your results, you may wish to export the results into a text file that you can examine in more detail. Select "Export Transcript" from the Transcript Menu, figure 3.50.

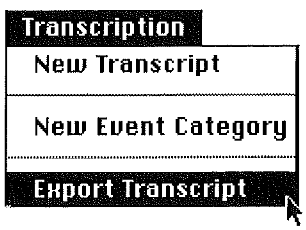


Figure 3.50. The Transcript Menu.

This will bring up a window asking you for the name of the file and where to save it. The resulting file is a text file that can be opened with any text editor or word processor. *Video Transcriptor* is not designed to analyse the results of the note-taking process. There are already many programs available to do exactly this. The results may be converted into the format needed for one of these specific statistical packages quite easily. Once the data has been imported into a statistical analysis package, a detailed analysis of the results can take place.

Menus

There are three menus in *Video Transcriptor*; the File, Edit, and Transcription menus.

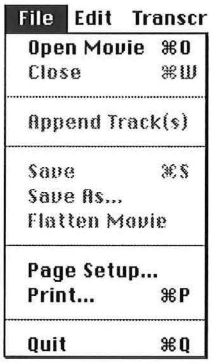
The *Video Transcriptor* menus list the commands available and their keyboard equivalents. To use a command, choose the command from its menu. To use the keyboard equivalent of the command, press the Command ⌘ key and type the letter shown.

About Video Transcriptor...

This menu option presents the user with the about box for *Video Transcriptor*. This contains version information about *Video Transcriptor* and information about the user's machine, figure 3.51.

Figure 3.51. About *Video Transcriptor*.

File Menu



The File Menu contains facilities related to operations on movie files. These operations include opening movie files, adding tracks to the current movie, saving the movie data and all its transcription data, and for printing the transcription data which has been entered.

Open Movie

In order to use *Video Transcriptor*, a movie which contains video and/or audio tracks must be opened. *Video Transcriptor* can only open movie files, it cannot create new movies, refer to the *Quick Start* Section. *Open Movie* brings up the open file dialog box for the user to select the movie which is to be analysed, figure 3.52.

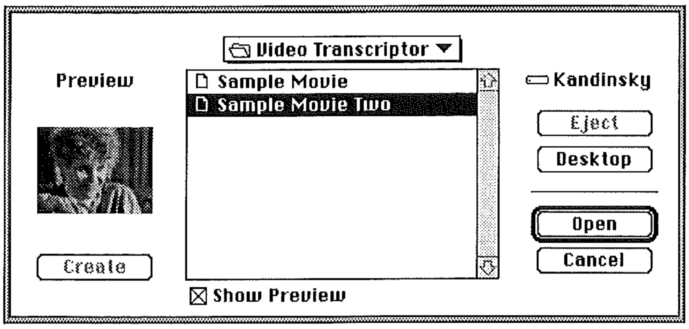


Figure 3.52. The Open File Dialog Box.

Only movie files are shown in the open dialog box. This makes it easy for the user to select valid movie files. The open file dialog box includes a preview area for still images and movies. This is not available in the conventional file dialog box but is an extended feature available as part of QuickTime.

Close

When the user is finished with a movie, but may not be ready to quit *Video Transcriptor,* the current movie may be closed by selecting *Close* in the File menu. If any changes have taken place since saving last time the user is prompted with a save file dialog box asking if the changes made are to be saved, discarded or if the user wishes to cancel the operation. Figure 3.53 shows the dialog box asking the user what to do with any changes.

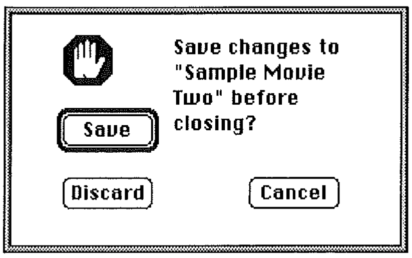


Figure 3.53. Close File Dialog Box.

The transcription data is saved into the current movie file if there are any changes needing saved. This command also closes the top most window.

Append Track(s)

The "Append Track(s) item is not available in the BASIC mode of *Video Transcriptor*. For more information on appending tracks refer to Section 5 - *Expert Mode Features* in this manual under "Menus".

Save

This option will save any changes made to the transcription notes in the current movie data. This option does not save the contents of the video and audio tracks into the movie file but uses pointers to the original copy of the movie. Instead of having two copies of the same video and audio information on the hard drive, pointers for some tracks in one movie may be pointing to another movie where they already exist. This process saves valuable disk space and is automatically handled by QuickTime. One important thing to note is that if the original movie file is thrown away then all other movies that rely on this movie file will be affected. To save the video and audio information with the new movie file refer to the save option *Flatten Movie* discussed shortly.

Save As...

The *Save As* menu option presents the user with the save file dialog box which can be seen in figure 3.54.

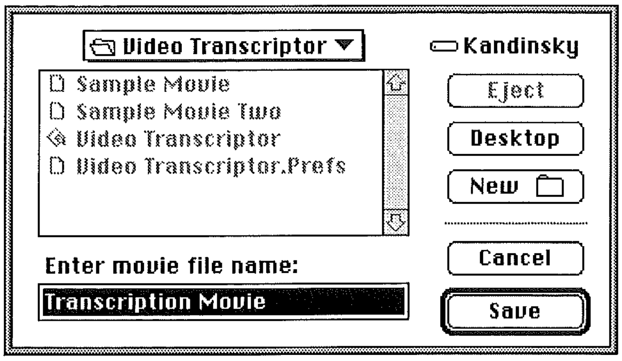


Figure 3.54. Save Dialog Box.

The user enters the new name for movie and its location on the hard drive. All changes which have occurred to the current movie are saved to this new movie file and the new movie becomes currently active. Any previous changes to the original movie are discarded and the original movie is closed.

Flatten Movie

The *Flatten Movie* option is similar to the *Save* option. It saves the entire movie data to disk. The difference with this is that all of the tracks and notes are saved into the one movie file. There are no pointers to other movie files where other tracks may reside as may occur with the *Save As...* option. The contents are actually copied to this new movie. This may potentially use more disk space but this option removes the problem of maintaining the pointers to individual tracks in different movie files. No loss of video or audio information will occur with this method.

Page Setup...

The *Page Setup* option brings up a dialog box for setting up page sizes and layouts for the screen and for printers. This can be seen in figure 3.55 below.

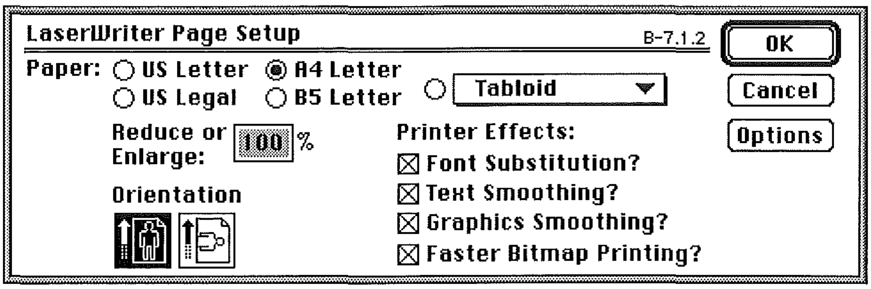


Figure 3.55. The Page Setup Dialog Box.

Print...

The *Print* option will print out a text format listing of all the transcription notes taken for the current movie. This listing ensures easy reference for the entered notes. Each track's notes are printed in order. Figure 3.56 shows the standard *Print* dialog box when printing the notes.

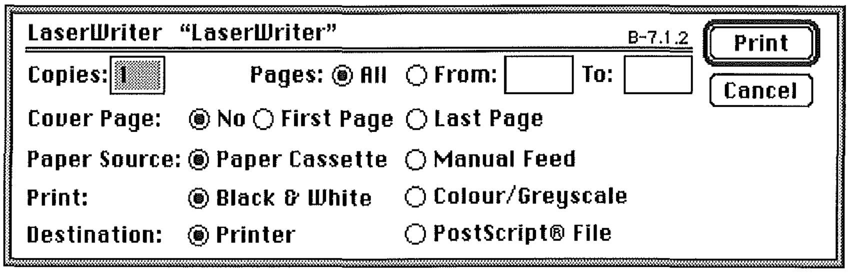
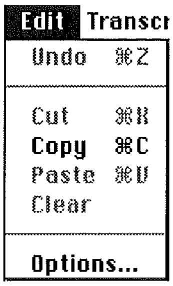


Figure 3.56. Print Dialog Box.

Quit

When the user chooses to *quit Video Transcriptor,* one of two possible options might occur. If no changes to the movie have taken place then the application exits. If there are changes which have occurred and have not already been saved, the user is prompted with a dialog box asking if the changes made are to be saved, or discarded or if the user wishes to cancel the *quit* operation. This is the same as mentioned above when closing a movie without saving changes that had been made. Any changes needing saved are recorded into the current movie file and then the application exits.

Edit Menu



The edit functions; undo, cut, copy, paste and clear, perform the standard Macintosh functions when applied to text in the note-taking section of *Video Transcriptor*. The *Options* menu item contains global settings for *Video Transcriptor*.

Undo

The *Undo* command cancels the last operation attempted. If an item was cut then it is returned to its previous location. If something was pasted then it is removed. This is very useful for when an unexpected event occurs.

Cut

This command removes the selected material and places it on the Clipboard; from there it can be pasted elsewhere.

Copy

This command copies the selected material and places it on the Clipboard, without removing the material from the original location. From the Clipboard it can be pasted elsewhere.

Paste

This command copies the contents of the Clipboard, whatever was last cut or copied, to the current insertion point.

Clear

This command clears the selected material from its location without placing a copy on the Clipboard. The *Undo* command can be used immediately after using Clear if a mistake was made.

Options

The *Options* menu item brings up the dialog box shown below in figure 3.57.

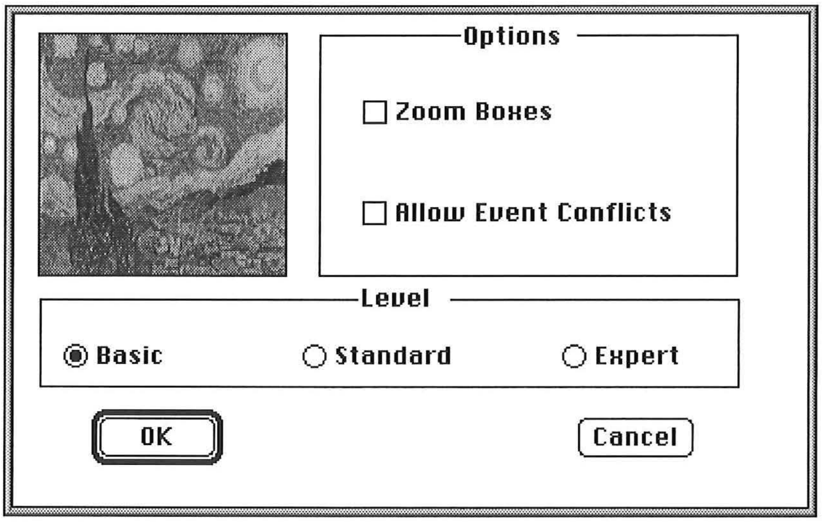


Figure 3.57. The Options Dialog Box.

The options dialog box allows the user to change global settings for *Video Transcriptor.* These settings include the mode, whether zoom boxes are available, and whether event conflicts are possible.

Mode

There are three modes available for note-taking; basic, standard and expert. Each mode allows more capabilities in the functionality of *Video Transcriptor.* More information about the different modes is available earlier in this section under *Mode Information.*

- **Basic**

The *basic* mode only allows the most simple elements of note-taking which are desirable by end-users. All this option is intended for is to allow the user to step through the movie using the standard control buttons taking the occasional notes. The positioning controls and speed controls are unavailable. The Zoom Boxes are turned off by default. The note-taking facilities are restricted by means of their functionality by only allowing one transcription window where the user enters notes for the entire movie in one place.

- **Standard**

The standard mode introduces the Zoom Box and the positioning controls back into the Master Controller. The Zoom Box can be turned off if desired independently of the *standard* mode. The note-taking facilities are improved by allowing a transcription widow for each track in the movie together with the transcription window for the entire movie.

- Expert

The expert mode contains all the available options turned on. The speed control is now available on the Master Controller for variable speed analysis of the movie data.

Zoom Boxes

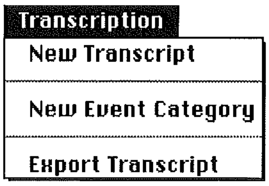
With *Zoom Boxes* option enabled, the ability to zoom into a video track is available. With this option turned off, the Zoom Box is removed from all displays. By default it is turned off in the *basic* mode. Only the most simple aspects of note-taking are desirable in this mode. The Zoom Box can be turned on or off independently of the mode which is currently set. The mode is only a default guideline for controls which should be available to the end-user.

Event Conflicts

The *Event Conflicts* option toggles between allowing multiple events to occur within the exact same time interval and restricting only one event entry within every time interval. The time interval being referred to is that of an individual frame.

Until now we have been operating at the BASIC level without zoom boxes and no event conflicts allowed. In the next section we will move onto the STANDARD level and discuss the new features available with this mode.

Transcription Menu



New Transcript

This command will activate the new transcription window for use in the note-taking process. Any notes taken in previous sessions will be loaded into the transcription window. There is only one transcription window in the BASIC mode of *Video Transcriptor* used for entering information about the movie as a whole. Individual transcription windows for each track are available in the advanced modes of *Video Transcriptor*. The functionality of this is discussed in Section 4 - *Standard Mode Features* in this manual under "Menus".

New Event Category

The *New Event Category* command is used in conjunction with a transcription window. It therefore assumes that there is a transcription window currently active. Each transcription window contains notes that can be divided up into categories. Each category contains notes of a similar nature. For example some categories when analysing people's behaviour might be; arm movements, leg movements, facial movements, facial expressions, eye movement and mouth movement. If notes were to be taken on all of these areas then it would be logical to break them up into these separate categories instead of grouping the notes altogether in one place.

When the *New Event Category* menu item is selected, a dialog box appears asking the user for the category name. The dialog box is shown in figure 3.58.

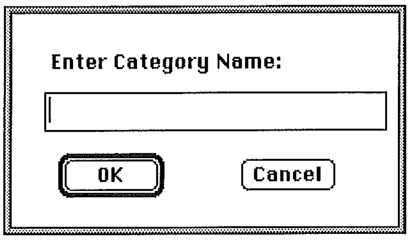


Figure 3.58. New Category Dialog Box.

The user enters a name and then presses *OK* to confirm the new category or presses *Cancel* to abort the process. Before any notes can be taken, an event category must be defined.

Export Transcript

Exporting the results from the transcription process is necessary before the analysis stage of the results can commence. The exported results can be imported into statistical analysis packages, databases and other programs. *Video Transcriptor* is not designed to analyse the results of the notes which are taken but is designed to provide a simple means for which the results can be exported for use in more specialised packages.

Upon selecting *Export Transcript* from the *Transcription Menu* the user is prompted with a save dialog box where the user can give the resulting file a name and determine where the file is to be saved to. The resulting file is a tab delimited text file. This option has been discussed earlier in this section under *Exporting Results*.

Section 4

Standard Mode Features

Introduction

This Section shows you how to use the additional features that are available with *Video Transcriptor* in the STANDARD mode. Once again this section is supposed to be a self guiding tutorial of the enhanced features available. Section 3 - *Tutorial* discusses the BASIC mode available and Section 5 - *Expert Mode Features* will discuss features available in the EXPERT mode.

Topics covered in this section:

- Changing Modes
- Zoom Control
- Position Controller
- Context Sensitive Transcription Windows
    - Track Component Method
    - Video and Audio Window Button Method
    - Menu Method
- Movie Time Resolution

- Changing the Time Resolution
    - Implications of Changing the Time Resolution
    - Multiple Events in One Time Resolution
- Transcription Process Enhancements
    - New Event
    - Modify Event
    - Remove Event
- Event Conflicts and Implications

This section is designed to complement Section 3 - *Tutorial.* It assumes that the user understands all of the features available in Section 3- *Tutorial.* Section 4 introduces more complex note-taking features based on each track in the movie instead of one transcription window for all tracks in the movie. The time resolution of each transcription window can be changed to assist in the note-taking process. This feature has many implications on how the note-taking process changes.

A zoom controller is available to assist in analysing the content of the video tracks in greater detail. The last major addition under the STANDARD mode is the positioning control. This gives instantaneous access to any location in the movie and greater control of playback.

We shall start this section by discussing how to change to the STANDARD mode and then the other new features of the Master Controller. Following this we will discuss the new features associated with the video and audio windows and then examine the implications facing the transcription window.

Changing Modes

At present we are still in the BASIC mode in *Video Transcriptor.* We can change the mode of *Video Transcriptor* in one of two ways; by the *Options* item in the Edit menu, or by clicking in the mode switch at the top of the Master Controller window.

Select *Options* from the Edit menu will bring up the options dialog box that contains global settings for *Video Transcriptor,* figure 4.1.

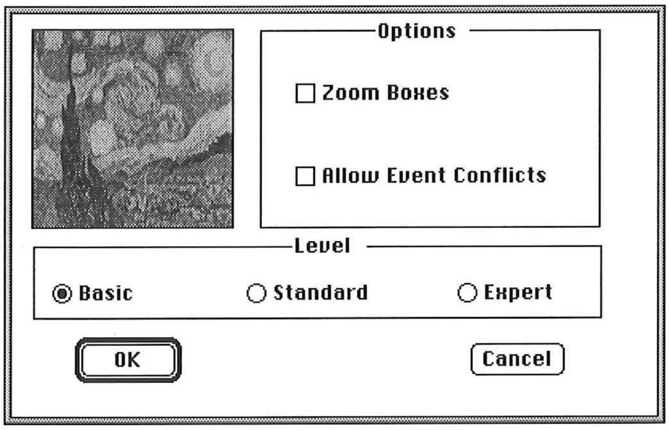


Figure 4.1. The Options Dialog Box.

Select 'Standard' as the new level and press the *OK* button. This will change the mode of *Video Transcriptor* to the STANDARD mode. For more information on the options menu item, refer to the menus section in Sections 3 - *Tutorial.*

You will notice that the "Zoom Controller" item is automatically enabled and the "Allow Event Conflicts" item is disabled. These items may be changed at any time regardless of the current mode. More information is available on these items later in this section. The Master Controller should now look like figure 4.2 below.

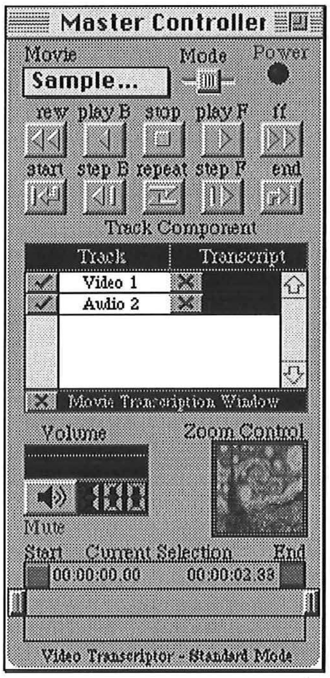


Figure 4.2. The Master Controller, STANDARD Mode.

The major visible differences on the Master Controller window is the zoom controller positioned to the right of the volume control and the positioning control at the bottom of the window. These new controls will be discussed next.

Zoom Controller

To the right of the volume control on the Master Controller is the Zoom control, figure 4.3.

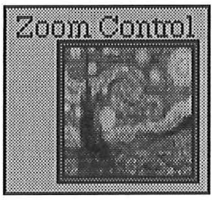


Figure 4.3. The Zoom Controller.

The purpose of the zoom control is to assist in analysing the contents of video windows in the movie. Its function is similar to enlarging a video window to see the contents better. The benefit with this control is that the video window stays the same size but an area of the contents is enlarged.

The picture represents a full-sized video window in the movie. Located around the edge of this picture is a resizable blue box. The contents of this box represents the portion of the whole movie that will be displayed in the video windows. If the box is the same size as the full box then the entire contents of the video information will be displayed. This is the default position when a new movie is loaded. As the box is resized, the area within the box is scaled to fit into the existing video windows.

By moving the mouse pointer over the edges of the blue box on the zoom controller, the mouse pointer changes shape indicating the possible mouse movement. For example when the mouse pointer is located over the left or right hand edges of the zoom controller's blue box the cursor changes shape into left and right pointing arrows, figure 4.4.

Figure 4.4. Left-Right Directional Movement.

Try pointing to the top or bottom of the blue box in the zoom controller. The mouse pointer changes shape into an up and down facing arrows, figure 4.5.

Figure 4.5. Up-Down Directional Movement.

This time try pointing to the corners of the blue box in the zoom controller. The mouse pointer changes shape into four directional arrows, figure 4.6.

Figure 4.6. Uni-Directional Movement.

Pressing the mouse button and dragging the mouse in the direction the arrows point resizes the zoom controller and resizes and the video windows content. Resize the zoom controller so it is one quarter of the original size by dragging the lower right hand corner to the centre of the box as shown in figure 4.7.

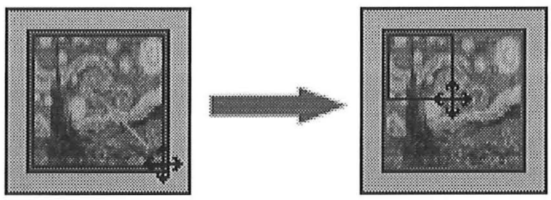


Figure 4.7. Resizing the Zoom Controller.

As you are resizing the zoom controller you will notice that the contents of the video window change. The portion of the video window that was in the upper left hand corner of the video window now takes up the entire contents of the window, figure 4.8.

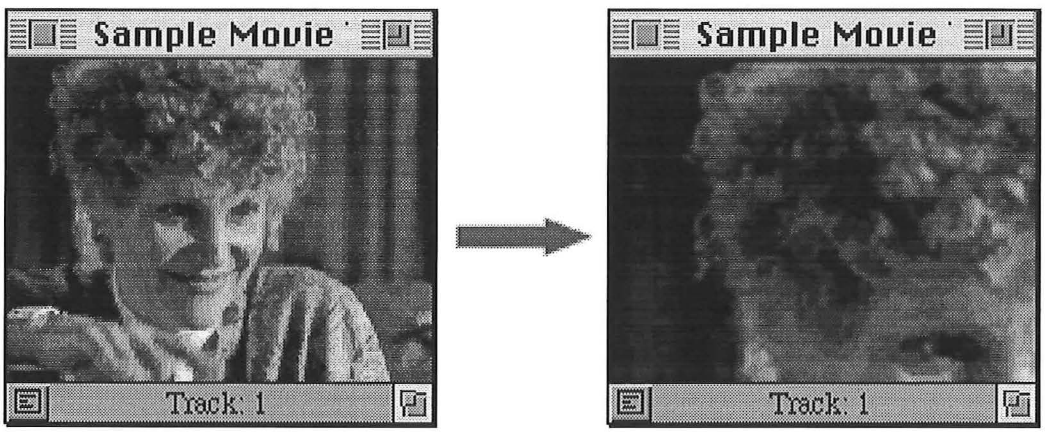


Figure 4.8. Effects of Resizing the Zoom Controller on the Video Windows.

If we map the zoom controller's location onto the video window then we can see the results we would expect to see, figure 4.9.

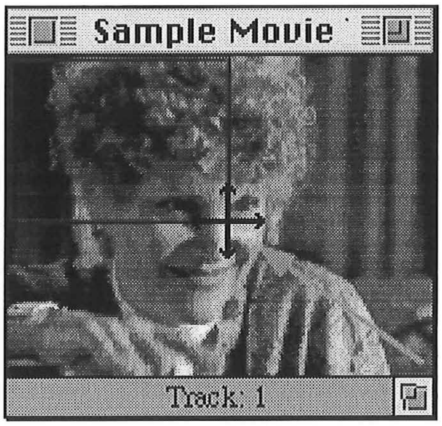


Figure 4.9. Mapping the Zoom Controller onto the Video Windows.

The area marked by the blue box is what we would expect to see of the video window after it has been scaled as we saw in figure 4.8.

Move the mouse into the centre of the blue box. The mouse pointer changes into an open hand, figure 4.10.

Figure 4.10. Mouse Pointer - Indicating the Ability to Movie the Zoom Controller.

Press down on the mouse button. The mouse pointer changes into a clenched fist, figure 4.11.

Figure 4.11. Mouse Pointer - Repositioning the Zoom Controller.

The resizable zoom box may be repositioned to another location within the movie poster box by dragging the mouse, figure 4.12.

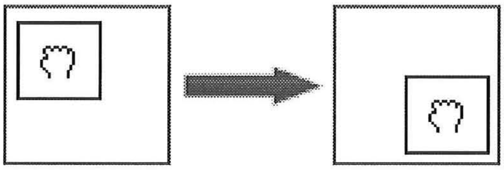


Figure 4.12. Repositioning the Zoom Controller Area.

You will see that the video window changes so that the lower right hand quarter of the video window is displayed.

Position Control

The lower portion of the Master Controller contains the Position Control, figure 4.13.

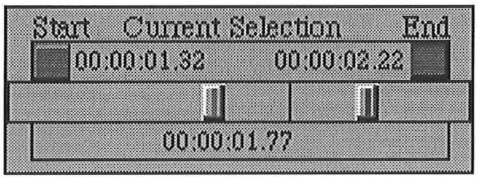


Figure 4.13. The Position Control.

The control is divided into three horizontal sections. The bottom section displays the current movie's time. The middle section is a graphical representation of the movie's current location, start, and end selection times with respect to the entire length of the movie. The uppermost section relates to the start and end time of the movie's active selection.

Current Movie Time

The bottom section of the Position Control displays the movie's current time. Press the "Play F" movie controller button to start the movie. As the movie plays watch the current movie time display. You will see that the display reflects the movie's current time. Press the "Stop" movie controller button and then the "Start" movie controller button to return to the beginning of the movie. The current movie time display should now read "00:00:00.00" indicating that the movie is now at the beginning.

Active Selection Controls

The middle section contains a graphical representation of the current active portion of the movie together with a representation of the current position of the movie with respect to the length of the entire movie.

The three indicators shown represent the start time (green coloured slider), the end time (red coloured slider), of the current active selection, and the movie's current time (black vertical bar).

By default the start indicator slider will be at the left hand edge at time zero and the end indicator slider will be at the right hand edge, the end of the movie time.

Press the "start" movie controller button to take the movie back to the beginning. Press the "play F" button. As the movie advances, watch the black bar move along to the right together with the current movie time mentioned previously. This black bar represents the movie's current location with respect to the length of the entire movie. Press the "stop" button.

The current active portion of the movie is the selected area that the user is working with. When the user presses the "play F" button on the Master Controller, the movie will only play between the start and the end times of the current active portion. Pressing the "repeat" button will force the current movie to loop over this section indefinitely. Pressing the "start" or "end" movie controller buttons will take the movie to the start or end of the current active portion of the movie respectively. The current active portion is a useful feature for examining a small segment of video or audio information in a movie. By setting the start and end times around the time to be examined, the user can utilise the standard control buttons more effectively. As mentioned, by turning the "repeat" function on, the movie will loop over and over this active portion of the movie. As a result of changing the current active portion of the movie, the start and end buttons for controlling the movie are more intelligent. Instead of going to the start and end of the movie, they will only relocate the movie to the start and end of the active portion.

Movie Indicator

Point to the movie's current time indicator (the black vertical bar) with the mouse. When the cursor is close to the black bar the mouse pointer changes shape and the current movie time display is highlighted in blue, figure 4.14 .

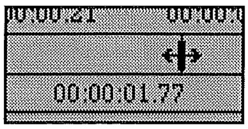


Figure 4.14. Change in Mouse Pointer.

Press the mouse button and drag the mouse to the left or to the right, the direction that the arrows point. As you drag the current time indicator you will notice that you are moving through the movie in the respective direction that you dragged the bar. While this is happening the video tracks are continuously being updated together with the current movie time display. If you have any transcription windows open, these will be updated also as you are moving through the movie. When you release the mouse button the movie now stops at the new location selected.

Try pointing to any location within the active movie area with the mouse and press the mouse button. The movie will immediately jump to this location and update all appropriate windows. This allows instantaneous access to any location within the movie. This is impossible to achieve under linear based technologies such as video-tape.

Start Indicator Slider

The start indicator slider is positioned at the left hand edge of the Position Control when a new movie is opened. When you point to the slider with the mouse you will notice that the current time of the indicator, initially located just above the indicator slider, changes colour to green together with the "Start" label above the upper section of the position control, figure 4.15.

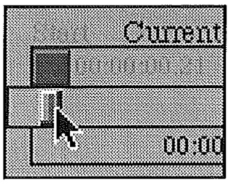


Figure 4.15. Start Indicator Slider Selected.

The highlighting makes it easy to see what has been selected by the user. Press the mouse button down and drag the start indicator slider to the right. The start time indicator display indicates the new time that is selected. You will notice that you cannot move the slider past the current time indicator. If you want it past the current time then you will have to move the current time indicator further along so that you can then move the start indicator slider.

Try pressing the "start" movie controller button. This will take you back to the position of the start indicator slider. It had previously taken you back to the start of the movie when you did this. You will also remember that this was where the start indicator slider was previously located. What the "start" and "end" buttons actually do is reposition the movie to the start or the end of the currently active portion of the movie and not to the "real" beginning or ending of the movie as it seems to in the BASIC mode and previous examples.

Press the "repeat" button and then the "play F' movie controller button. You will see that when the movie reaches the end of the active portion of the movie it will loop back to the start indicator slider's position and keep on playing. Press the "stop" button to stop the move and press the "repeat" button again to toggle the looping function off.

End Indicator Slider

The end indicator slider acts the same as the start indicator slider. When selected, the end time indicator is highlighted in red, figure 4.16.

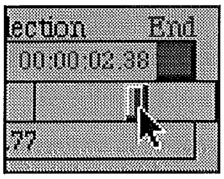


Figure 4.16. End Indicator Slider Selected.

Like the start indicator, the end indicator cannot be moved beyond the current time indicator. The current time indicator must be moved first. All of the movie controls depend on the active movie selection. The movie can only be positioned within this active time range. The start and end times can be repositioned at any time that is required.

Movie Selection Controls

The uppermost section of the position control contains the movie selection controls. This consists of two buttons and two time displays. The two movie selection buttons, labelled "Start" and "End", will set the start and end indicator slider times to that of the current movie time. On a colour display the green left-hand button sets the start time and the red right-hand button sets the end time. Beside each of these buttons is a display that indicates the start and end times of the active movie selection relative to the start of the movie. Position the movie to the beginning of a section of the movie that you would like to examine. Press the "Start" movie selection button. You will notice that the start time beside the start button has changed. This time is now the same as the current time of the movie. When you pressed the "Start" movie selection button the start indicator slider moved automatically to its new location and the start time display was updated.

Reposition the movie to the end of the movie section that you wish to examine. Press the "End" button. The end indicator slider is repositioned to the current movie time and the end time display beside the "End" button is updated.

You have now made this section of the movie the 'active' section of the movie. Try pressing the "start" movie controller button, the button near the top of the Master Controller. You will notice that the movie is repositioned to the time of the start indicator slider and not the actual beginning of the movie. Previously in the BASIC mode the whole movie acted as the active selection so when the "start" movie controller button was pressed you were always taken back to the start of the movie. Likewise the "end" movie controller button will take the movie to the end indicator slider position and not the end of the movie. Press the "start" and then the "play F" movie controller buttons. You will see that only the active selection of the movie is played.

Press the "start" and then the "repeat" movie controller button followed by the "play F" button. When the movie gets to the end of the active movie selection this time it loops back to the start of the active selection indefinitely until you press the "stop" movie controller button. Press the "repeat" movie controller button to toggle the looping function off.

The active movie selection facility is useful when you want to examine a small segment of video or audio information from the entire movie in detail.

Context Sensitive Transcription Windows

The STANDARD mode of *Video Transcriptor* offers more flexibility in the area of note-taking than the BASIC mode. In Section 3 - *Tutorial* we saw that *Video Transcriptor* offered a single transcription window for note-taking. Under the STANDARD mode, *Video Transcriptor* offers multiple transcription windows; one for each audio and video track in the movie. The effects of this are discussed shortly. The STANDARD mode also offers the ability to change the time resolution of the note-taking allowing the user to define time periods for notes to be taken within.

There are now three ways to obtain a transcription window. These are; from the Track Component on the Master Controller, a button on the video and audio windows, and from the Transcription menu.

Track Component Method

The functionality of the Track Component has already been discussed in Section 3 - *Tutorial*. The STANDARD mode in *Video Transcriptor* offers additional features for controlling the transcription windows in the form of a second column labelled "Transcript", figure 4.17.

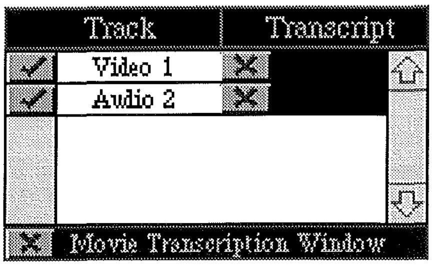


Figure 4.17. Track Component, STANDARD Mode.

All of the existing features discussed on the Track Component behave exactly as they did in the BASIC mode.

The second column of the Track Component controls the transcription windows used in the note-taking.

Pressing a button in this column will open or close a transcription window for the particular track. A cross on the button means that there is no transcription window open on the desktop and pressing it will open one. A tick on the other hand means that there is a transcription window already open. Pressing the button in this situation will close the window. Each of these transcript windows apply to only one video or audio track in the movie and consequently all notes taken in a particular transcription window only apply to the associated track.

When you are finished with a transcription window you may either close the window by clicking in its close box or by pressing on the button for the transcription window.

Another useful feature for handling the transcription windows is double-clicking on the black area located just to the right of one of the transcription window buttons. If the transcription window is open but many other windows are on top of it, double-clicking in this black area will bring the transcription window to the front. If the window is not already open then it will open the window just as it would if you had pressed on the corresponding button in the "Transcript" column.

Video and Audio Window Method

The video and audio windows have also changed under the STANDARD mode of *Video Transcriptor*. Both windows now have an additional button in the lower left hand corner, figure 4.18 for the video window and figure 4.19 for the audio window. This button is the "Transcription" button.

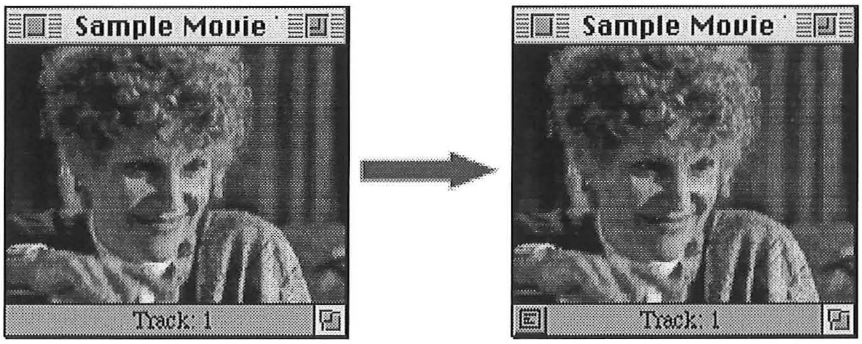


Figure 4.18. Video Window, BASIC and STANDARD Modes.

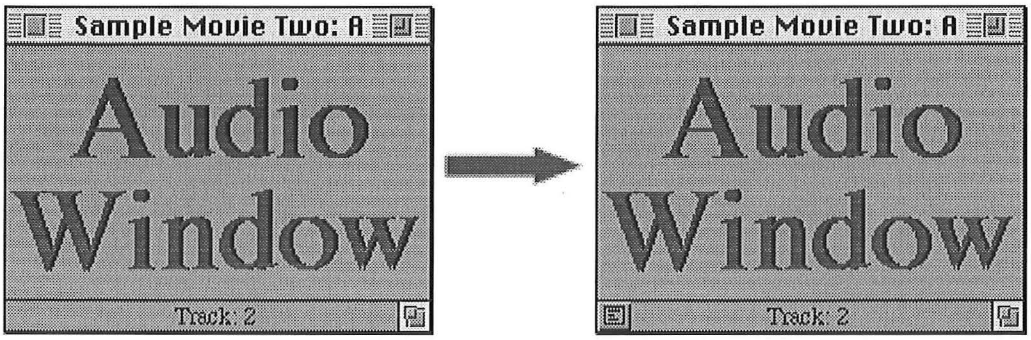


Figure 4.19. Audio Window, BASIC and STANDARD Modes.

This "Transcription" button will activate the transcription window for the current video or audio track. Any notes already taken will be shown. If the window is already open then it will be brought to the front.

Context Sensitive Transcription Menus

The only means of obtaining a transcription window in the BASIC mode was through the *New Transcript* menu command in the Transcript menu. This opened a single transcription window for the whole movie. With the ability to have multiple transcription windows, one for each video and audio track in the movie, the *New Transcript* command has become sensitive to the currently active track in the movie.

To open the transcription window for a video or an audio track the video or audio track window must be open. Click in the window to make it the front most active window. Select *New Transcript* from the Transcript menu. This will open the appropriate transcription window for the track.

You can also open the original single transcription window for the whole movie just as you did in the Section 3 - *Tutorial*. This is the same transcription window that is used in the BASIC mode for the note-taking process. It is used for notes that affect the movie as a whole and not an individual track. Any notes that may have been entered in the BASIC mode can be seen here. To open the transcription window for the whole movie, select the Master Controller window and select the *New Transcript* menu item from the Transcription menu.

Movie Time Resolution

The STANDARD mode in *Video Transcriptor* offers the ability to change the resolution of the time interval for each transcription window. This allows the user to define intervals for notes. When analysing video data the researcher may want to take notes in five second intervals. By changing the track's time resolution to a five second interval each five second interval will appear as a single entry. Each interval, as a result of this, may contain multiple event entries. The user still enters the events at the level of a single frame in the track, not at the level of the whole interval.

For example if we change the resolution to five seconds and an event occurred at time six seconds, we would enter the details at the time six seconds, not as an event for the period of five to ten seconds.

The time resolution component of the transcription window is located on the left hand side of the window, figure 4.20.

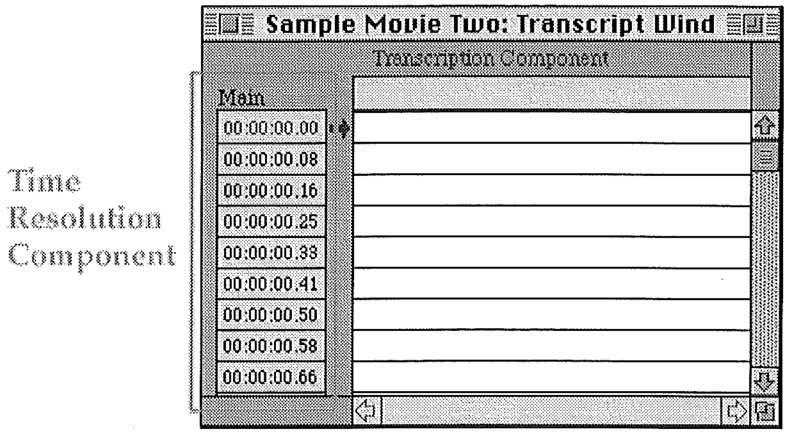


Figure 4.20. Transcription Window, Time Resolution Component.

By default the time resolution for a new transcription window is the length of a single video frame being played back at normal speed, figure 4.21.

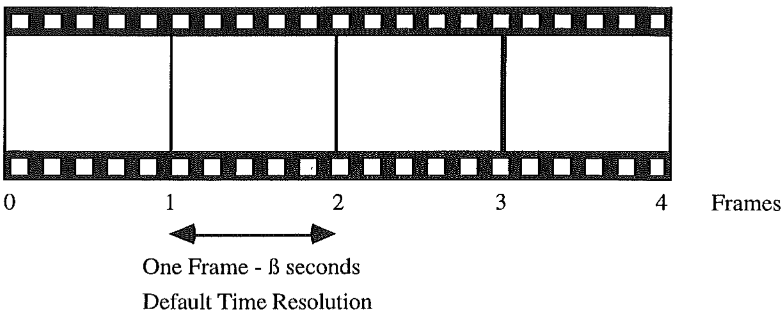


Figure 4.21. Default Time Resolution.

If we have fifteen frames being displayed in a second then the length of a single video frame would be $\frac{1}{15}^{\text{th}}$ second. The time resolution cannot be less than this but it may be changed to any value up to the length of the movie. The exact value will be changed to be a multiple of the default time resolution, the time taken to display one video frame.

Each second of movie data has multiple still-images which are shown per second. Each of these still-images are referred to as a frame. The number of frames which are shown per second is known as the movie's frame rate. By default the movie's time resolution is the same time it takes to show a single frame. In figure 4.21, the time resolution would be set to ß seconds, where ß is the time to show a single frame at normal playback speed.

For example if the source data has a frame rate of fifteen frames per second then the time resolution is $\frac{1}{15}$th seconds. The time resolution can not be less than the duration of one frame. It can be increased to any size but it will be rounded to become a multiple of the duration of one frame.

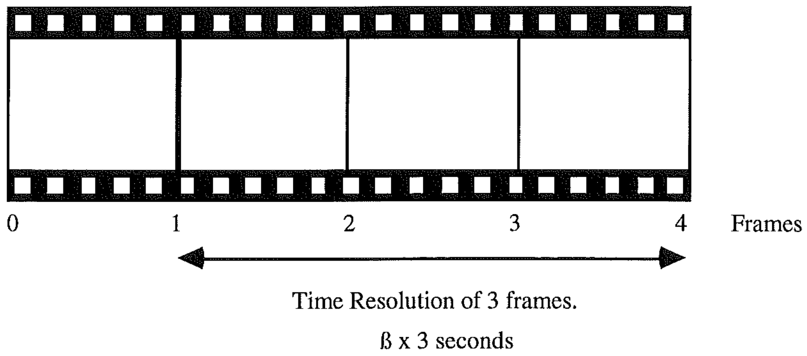


Figure 4.22. Changing the Default Time Resolution.

The default time resolution of one frame can be changed to assist in better time intervals for note-taking. In Figure 4.22, the time resolution is changed from being the time interval of one frame to that of three frames. The time resolution changes to ß x 3 seconds, where ß is the time it takes for one frame to be displayed. In practice situations the user may want to change the time resolution into manageable units like half seconds or seconds. The user doesn't need to know how many frames this will be but just the time resolution required. The time resolution suggested by the user is adjusted to align with an exact number of frames which will pass during this time interval.

Changing the Time Resolution

To change the time resolution of the transcript window, the user clicks the mouse button in the time resolution display area.

Any changes made to the time resolution will only affect the current transcription window. Figure 4.23 shows the time resolution dialog box that appears for the user to change the time resolution of the transcription window.

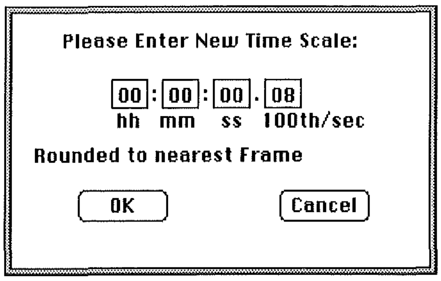


Figure 4.23. Dialog Box for Changing the Time Resolution.

The time resolution consists of an entry for hours, minutes, seconds, and hundredths of a second. The user uses the mouse to select the appropriate field to be changed and types in the new value. When the user is finished the *OK* or *Cancel* button can be pressed. The default time that appears is the time of one frame rounded to the nearest hundredth of a second. In figure 4.23 the time resolution is 0.08 seconds for each time interval. To change the time resolution for the transcriptions enter the new time into the time resolution dialog box. Enter 0.50 seconds as the new time resolution for the track, i.e. enter "50" into the 100^{th}/sec field replacing the "08" that was there previously.

Each frame in a movie has a very accurate time resolution. *Video Transcriptor* only represents this value down to the hundredth of a second scale. In our example our time resolution is 0.08 seconds but in reality it is slightly more than this but is not represented to this accuracy. We just changed the time resolution to 0.50 seconds when we would expect the time resolution to be accurate at 0.48 . Over this time interval there is enough overhead in the actual time resolution for an additional 0.02 seconds of time. The overhead in each frame is around 0.003 seconds per frame. So every three frames yields an extra hundredth of a second onto the time.

When a new time resolution has been entered, the time resolution is rounded to the nearest hundredth of a second. This rounding ensures that all events entered during the note-taking process are recorded properly. Problems could occur if the time resolution was changed frequently, shifting events from what was previously associated with one frame to another. Without this strict enforcement of having time resolution intervals being aligned with frame boundaries, event entries could shift around leaving data in an unstable state to be accurate. With the time resolution being adjusted automatically, all problems of this nature are eliminated.

If an error is made when changing the time resolution, the user is notified by an alert. Figure 4.24 shows an error caused by entering an invalid entry into the "seconds" field. Similar alerts also appear for other invalid entries.

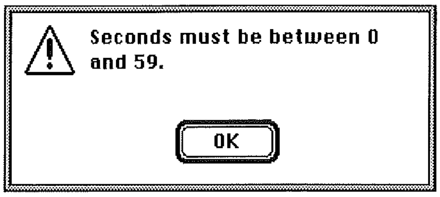


Figure 4.24. Error in Entering a New Time Resolution.

Implications of Changing the Time Resolution

By changing the time resolution from its default setting, the implications are apparent in the transcription window. The time resolution display area immediately changes to reflect any changes made to the track's time resolution. The default display area showed intervals of 0.08 seconds in each row when we first opened the transcription window. When we changed this resolution to 0.50 seconds we could see that the time intervals in each row changed to 0.50 second intervals. The contents of the event cells also changed. Previously the short description of each event was displayed in the event cell. Since changing the time resolution, the contents of the visible cells also changed.

Multiple Events in One Time Cell

Since the time resolution changed from 0.08 seconds to 0.50 seconds the format of each event cell changed. The default resolution, as in the BASIC mode, only allowed one event per time interval, figure 4.25.

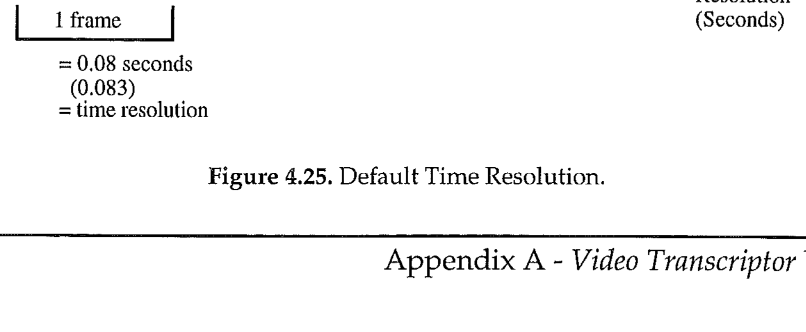


Figure 4.25. Default Time Resolution.

Since we changed the time interval to 0.50 seconds, we can now potentially have events occurring during one time interval, figure 4.26.

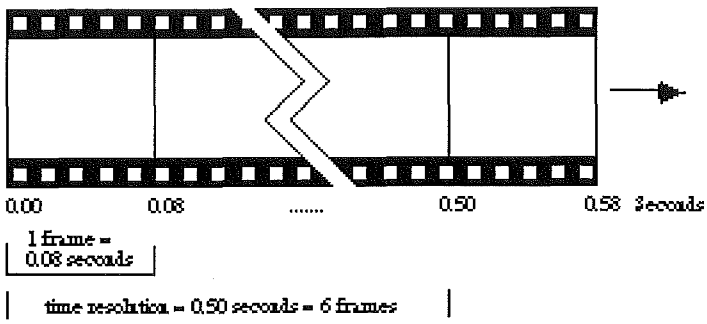


Figure 4.26. Changed Time Resolution.

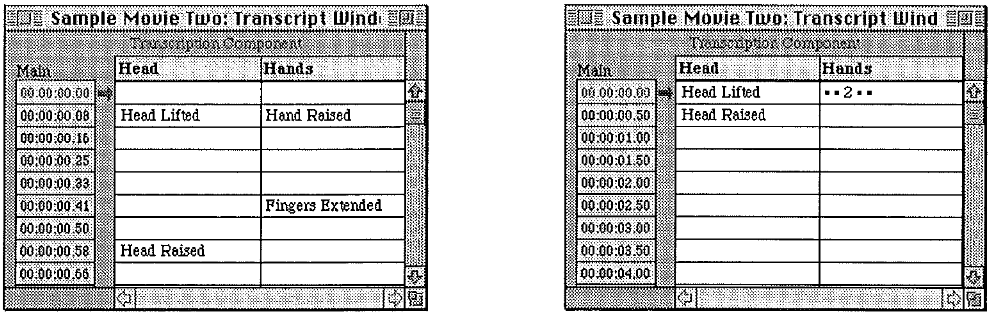
In the transcription window we normally display the short description field about each event in the appropriate event cell. This is normally fine when in the BASIC mode or the time resolution has not been changed. If the time resolution has been changed, as in our case of changing it from 0.08 to 0.50 seconds per time interval, we may have one of two situations for each event cell. The first situation is when we have zero or one event that occurs within a single time interval, the second situation is when two or more events occur during the same time interval.

Zero or One Events in One Time Interval

When zero or one events occur during a time interval, then it is displayed as it was before changing the time resolution for the track. In the case of zero events there was nothing displayed previously and nothing is displayed now in a cell. Where there was one event, the same event details are displayed.

Two or More Events in One Time Interval

When two or more events occur within the same time interval we cannot display the information about all of the events in the cell in the transcription window. Instead an entry is displayed that indicates that multiple events occur during the one time interval. the format of this entry is "• • # • •" where the "#" character is the number of events that occur during this interval. For example "• • 2 • •" would indicate that two events occurred during the time interval while an entry "Head Lifted" would indicate that it was the only event during the time interval. Figure 4.27 shows an example of changing the time resolution from 0.08 to 0.50 seconds on events in the transcription window.



Default Time Resolution (0.08 seconds). Changed Time Resolution (0.50 seconds).

Figure 4.27. The Effects of Changing the Time Resolution on the Event Cells.

We can see in the "Head" event category that only one event occurs in the first 0.5 seconds. Looking at the transcription window after we changed the resolution we can still see the single event entry. The same applies in the second time interval, 0.5 - 1.0 second where another single event occurs. During the first half second period in the "Hands" event category we can see that we have two events. After changing the time resolution we have an entry "• • 2 • •" that indicates two events occurred during the time interval. During the next half second period, 0.5 - 1.0 second, we can see that no events occur. After we changed the time resolution we have a blank entry during this time interval. This reflects that no events occur during this time interval.

Transcription Process Enhancements

The STANDARD mode of *Video Transcriptor* has brought many new capabilities to the note-taking process. Each track within the QuickTime format movie may now have a transcription window for you to take notes in. We have also seen the other major change to the note-taking process; the ability to change the default time resolution for the transcription window. This ability to change the time resolution has many consequences to the note-taking process of *Video Transcriptor.*

We shall examine the changes in the three categories; new events, modifying events, and removing events. We will see how the process of note-taking has changed since we have changed the time resolution to 0.5 seconds from a 0.08 second time interval for the main transcription window. The rest of Section 4 - *Standard Mode Features* will refer to the main transcription window with the modified time resolution.

Everything that applies to the creation, modification, and removal of event categories in the BASIC mode of operation still applies in all modes of *Video Transcriptor.*

Creating a New Event

Before we begin we will examine what we should see in the main transcription window at this point, figure 4.28.

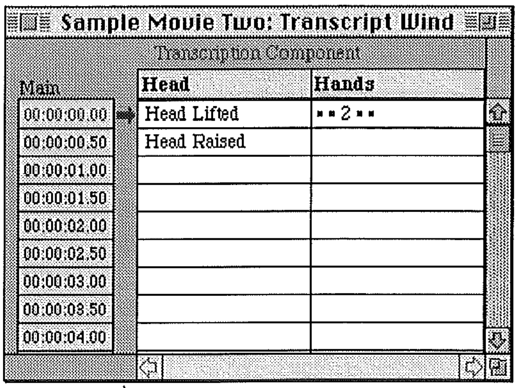


Figure 4.28. Main Transcription Window

We only have four events recorded in the transcription window. We will create new events under certain conditions in the transcription window to show the different possibilities presented to a user. There are two main ways to create a new event; by double-clicking on an event cell, or by the pop-up menu presented by holding the mouse button down while pointing to an event cell.

Position the movie to the location 0.66 seconds. This is in the second time interval of our track. Under the BASIC mode if we double-clicked in an empty cell we would instantly be able to enter information about the new event.

If we double-click the mouse button while pointing in the "Hands" event category at the time interval labelled "00:00:00.50". You will be confronted with a window for selecting the time that you wish to enter the event into, figure 4.29.

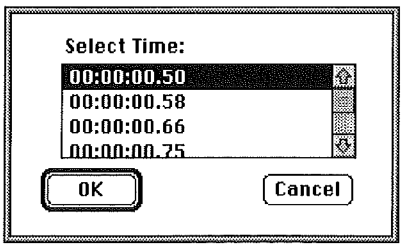


Figure 4.29. Selecting a Time For a New Event.

This list contains an entry for each vacant event entry within the time interval. Select the entry "00:00:00:.66" and then press the *OK* button. The window for entering event information will then appear, figure 4.30.

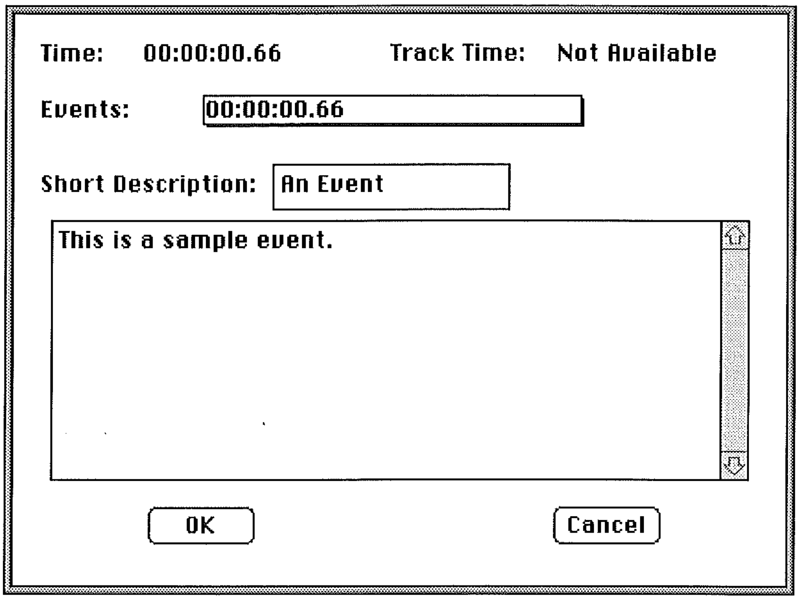


Figure 4.30. Window for Entering New Event Information.

Enter "An Event" into the short description field and "This is a sample event" into the long description field. The transcription window should now look like figure 4.31.

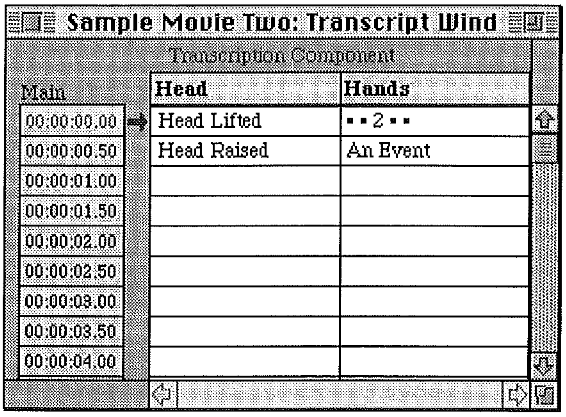


Figure 4.31. Transcription Window With New Event "An Event".

Reposition the movie to time "00:00:00.16" seconds and then double-click on the event cell in the "Hands" event category. There are already two events present during this time interval. Instead of getting the same window as in figure 4.29 to select a time, a different window appears, figure 4.32.

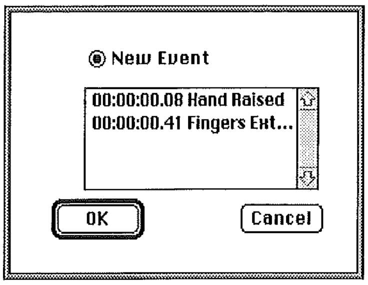


Figure 4.32. Event Time Selection Window.

This window asks whether you wish to create a new event or modify one of the two existing events during this time interval. By default the "New Event" item is selected. Press the *OK* button. A second window appears asking for the time of the event that you wish to create, figure 4.33.

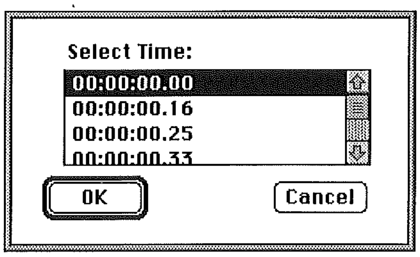


Figure 4.33. Selecting the Time For the New Event.

You will notice that the time "00:00:00.08" is not in the list. This time has already been used by the "Hand Raised" event. The list only consists of available times for selection. Select the time "00:00:00.16" and press the *OK* button. The window for entering new event information appears, figure 4.30. This time enter "Another Event" into the short description field and "Another sample event" into the long description field. When you have done this the main transcription window should now look like figure 4.34.

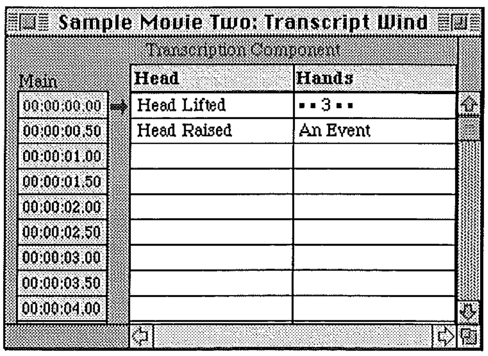


Figure 4.34. Main Transcription Window.

When you have entered the maximum number of events that you are allowed during the time interval then you are unable to enter any new events. Since we changed the default time resolution of 0.08 seconds to 0.50 seconds, the maximum number of events that we can have is six. So after we have created six events we have the maximum number of events during the time interval. If we try to create another event we will get a different window appearing such as figure 4.35.

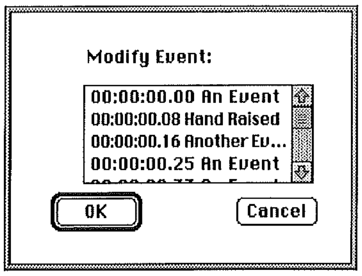


Figure 4.35. Selection Window For Modifying an Event.

This window only allows us to modify one of the six events. To modify an event just select the event and then press the *OK* button otherwise press the *Cancel* button to abort this process. This is only a demonstration of what will happen. Modifying events will be discussed next.

Creating a New Event Via The Pop-up Menu

As we saw in Section 3 - *Tutorial* we can also use the pop-up menu to create a new event. Point to the time interval labelled "00:00:01.00" seconds in the "Hands" event category and hold the mouse button down. This time interval contains no events. A pop-up menu will appear, figure 4.36.

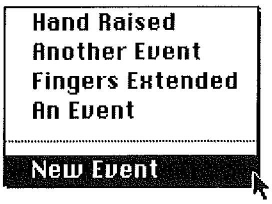


Figure 4.36. New Event Pop-up Menu - No Events in Time Interval.

You will see that the pop-up is divided into two; the top section contains a list of previous event titles, the bottom section will allow us to create a new event. More details of this process can be found in Section 3 - *Tutorial* under "Creating a New Event".

Click in the time interval labelled "00:00:00.00" in the "Hands" event category to bring up the pop-up menu, figure 4.37.

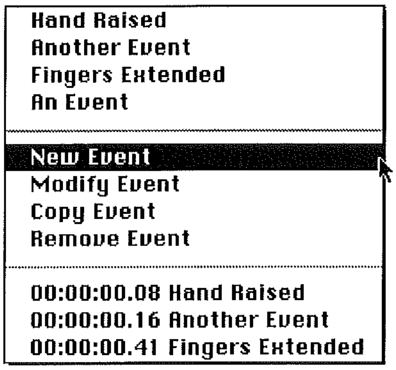


Figure 4.37. New Event Pop-up Menu - Events in Time Interval.

This presents us with a large pop-up menu that is divided into three sections. The top section contains a list of event titles that have already been inserted into the "Hands" event category. Selecting one of these will bring up a window asking the time to insert the event, figure 4.33, and then enter it automatically. The middle section contains actions that can be performed on the various events within the time interval. The bottom section contains a list of all events that have been entered within the current time interval. Selecting one of these will bring up a window to change the contents, figure 4.30 for example.

There are numerous combinations of menu options that depend on several factors:

- Are there any events already defined in the event category?
- Are there any events already defined in the current time interval?
- Is there an event in the special clipboard buffer?
- Is there an event already defined at the current time?
- Has the maximum number of events per time interval been exceeded?

Based on these different conditions we get different menu options. At most there can be three sections to the pop-up menu. The top section usually consists of a list of previously defined events that can be used to automatically generate a new one. The middle section contains a list of commands that depend on the current context of the note-taking. This may have:

- New Event
- Modify Event
- Copy Event
- Paste Event
- Remove Event

These menu options only apply to the event location at the current time. The bottom section contains a list of events that have occurred. By selecting one of these you may modify the contents of that event.

Modifying an Event

Modifying an event can be accomplished in one of two ways; by double-clicking in a time interval, or by the pop-up menu. Double-clicking the time interval brings up the window to enter a new event or modify an existing one, figure 4.38.

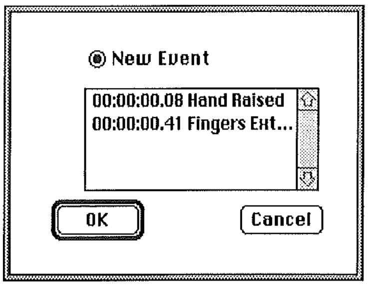


Figure 4.38. Event Time Selection Window.

If the time interval is full of events then a different window will appear, figure 4.39.

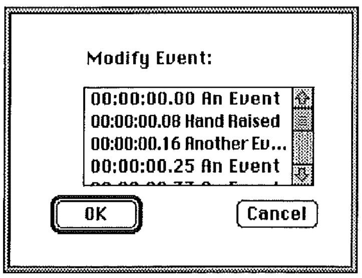


Figure 4.39. Selection Window For Modifying an Event.

In either window an existing event can be selected and then press the *OK* button. This will bring up the window to modify an existing event, figure 4.30.

By selecting *Modify Event* from the pop-up menu item, the same window as figure 4.39 for selecting an event to modify is opened. The user skips over the step of deciding whether or not it is to be a new event or whether an existing event is to be modified by selecting to use the pop-up menu method.

Removing an Event

Removing an event can only be achieved by selecting it from the pop-up menu method. This brings up a window to select an event to be removed, figure 4.40.

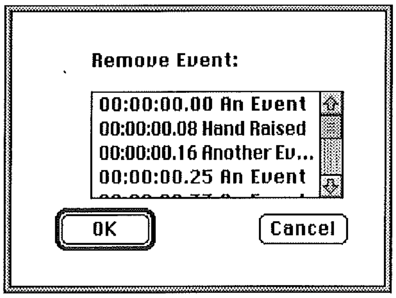


Figure 4.40. Selection Window For Removing an Event.

Once an event is selected press the *OK* button. The event is removed from the transcription window and copied to the special clipboard buffer.

Pasting an Event

Selecting *Paste Event* from the pop-up menu brings up a window for the user to select a time for insertion into the time interval, figure 4.41.

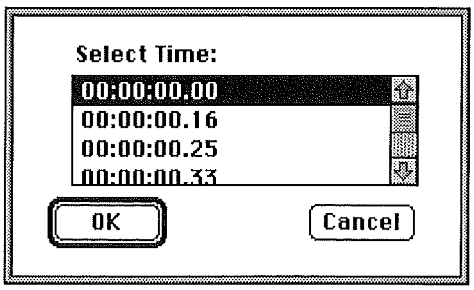


Figure 4.41. Selecting the Time For Pasting a New Event.

Copying an Event

Selecting *Copy Event* from the pop-up menu brings up a window for the user to select the time of the event to be copied to the special clipboard buffer, figure 4.42.

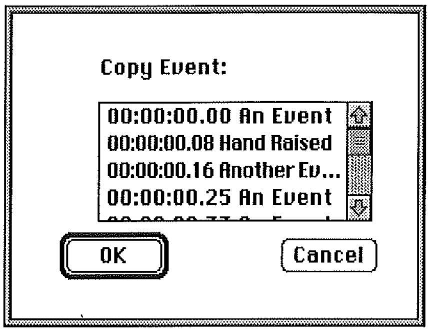


Figure 4.42. Selection Window For Copying an Event.

Pop-up Menu States

The following illustrates the various states of the pop-up menus in the STANDARD and EXPERT modes of *Video Transcriptor*. Figure 4.43 shows the differences in the basic principles between having one event per time interval and potentially many per time interval when double-clicking the mouse button in a time interval.

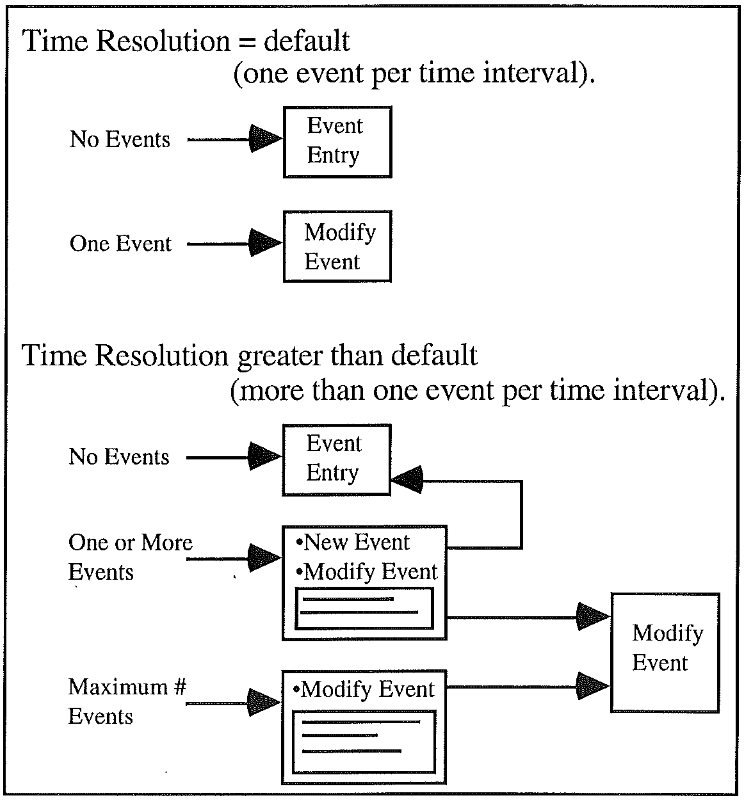


Figure 4.43. Comparison of the Procedures for the Creation and Modification of Events With One or Many Events Per Time Interval.

Figure 4.44 shows all of the possible menu combinations that may occur, depending on which conditions apply that were discussed previously.

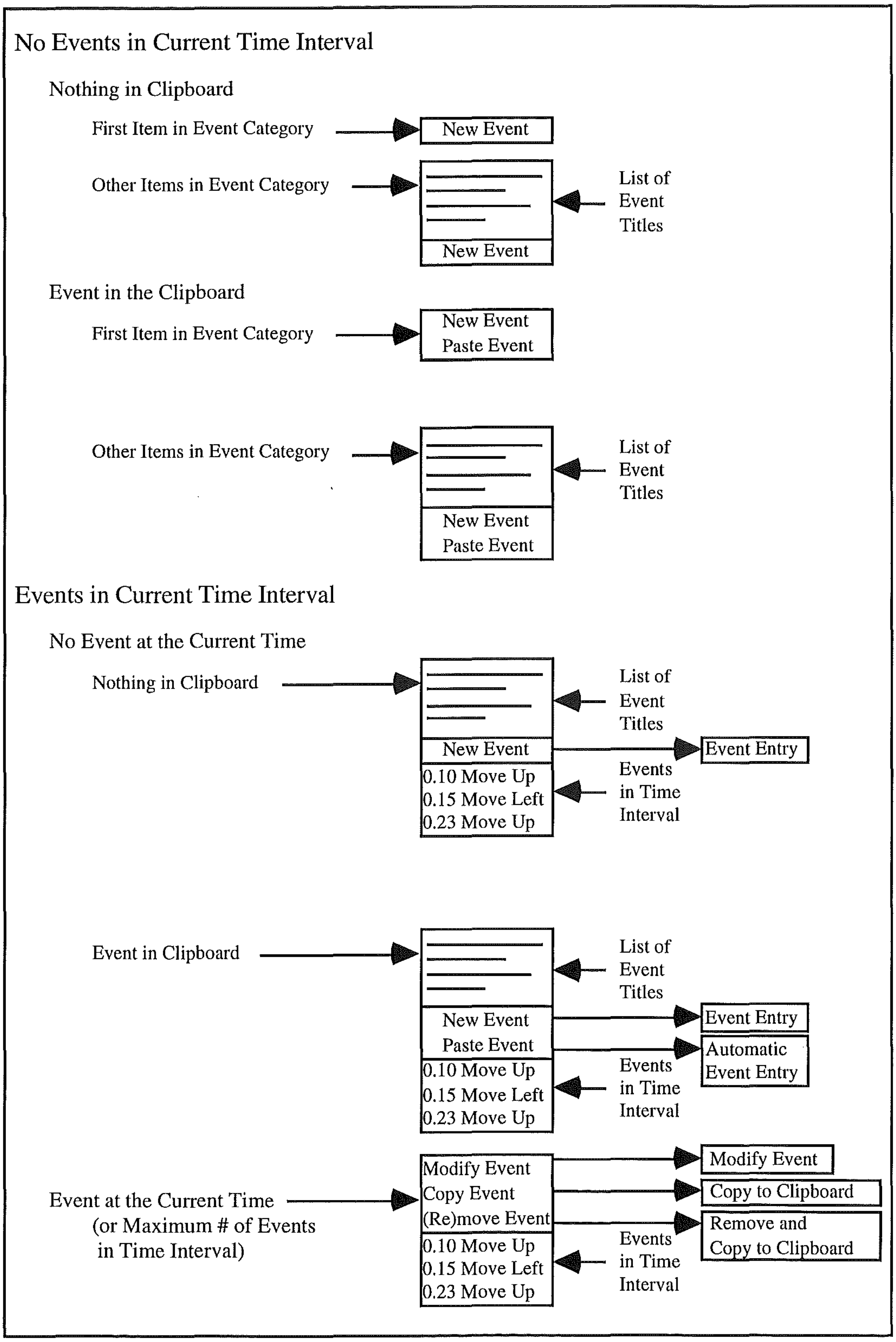


Figure 4.44. Possible Menu Options.

The results of these menus are:

- Event Entry

  A window for entry about a new event appears, figure 4.45.

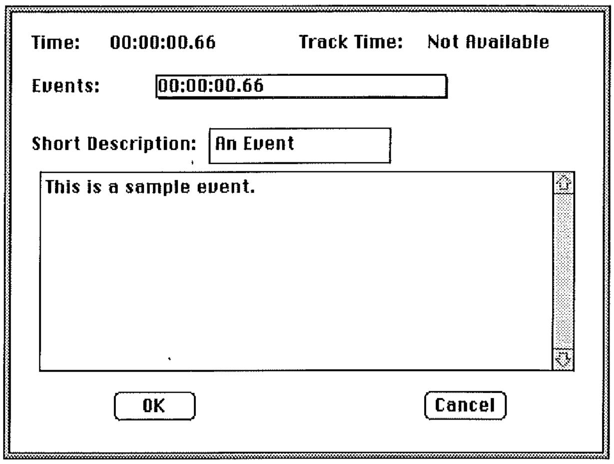


Figure 4.45. New Event Entry Window.

- Modify Event

  The same window as in figure 4.45 appears but with the current event details already filled in. There must be an event at the current time.

- Automatic Event Entry

  The short description field selected is automatically inserted into a newly created event.

- Copy Event to Clipboard

  The event details are copied to memory so they may be inserted elsewhere. There must be an event at the current time.

- Remove an Event

  This copies the event details to memory but also removes the original event. There must be an event at the current time.

More information about these items may be found in Section 3 - *Tutorial,* under "Transcription Window".

Event Conflicts and Implications

The last major change in the STANDARD mode in *Video Transcriptor* is the ability to have event conflicts. This means that it is possible to have multiple events that can occur at the same time and in the same event category. By default this option is disabled and should be used wisely. To activate it select *Options* menu item from the Edit Menu. Within the window that appears, figure 4.46, there is a check box to enable this option.

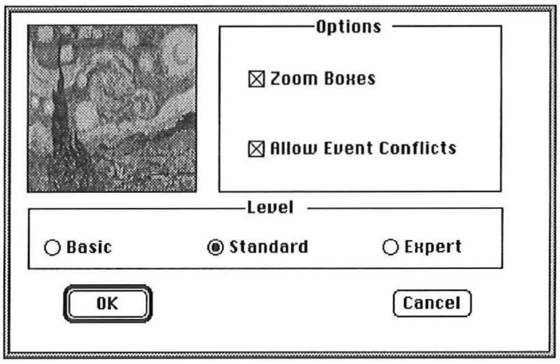


Figure 4.46. The Options Window.

If you start to use conflicting event entries you will be unable to read any extra events that you enter if you decide to turn off the event conflicts option. The actual events are not actually lost but are unavailable until you switch back to having event conflicts again. By default you can only have one event occurring at anyone time in each event category. Figure 4.47 illustrates the differences of the event conflicts.

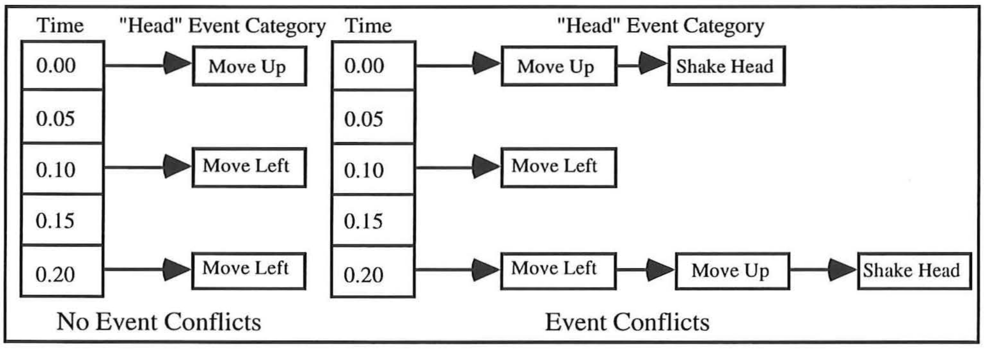


Figure 4.47. Events Occurring at the Same Time, (Event Conflicts).

The process of creating and modifying events changes with "Event Conflicts" turned on. There is no longer a maximum number of events that can occur within a time interval. Figure 4.48 illustrates what occurs when double-clicking the mouse to create or modify events when the "Event Conflicts" option is turned on.

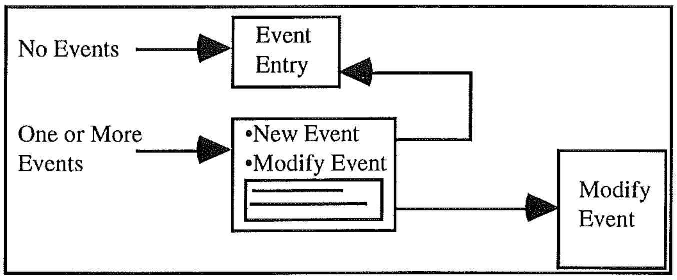


Figure 4.48. Effect of Event Conflicts on Creating and Modifying Events.

Pop-up menus are also affected by the introduction of "Event Conflicts". Figure 4.49 shows the possible menu options with "Event Conflicts" turned on.

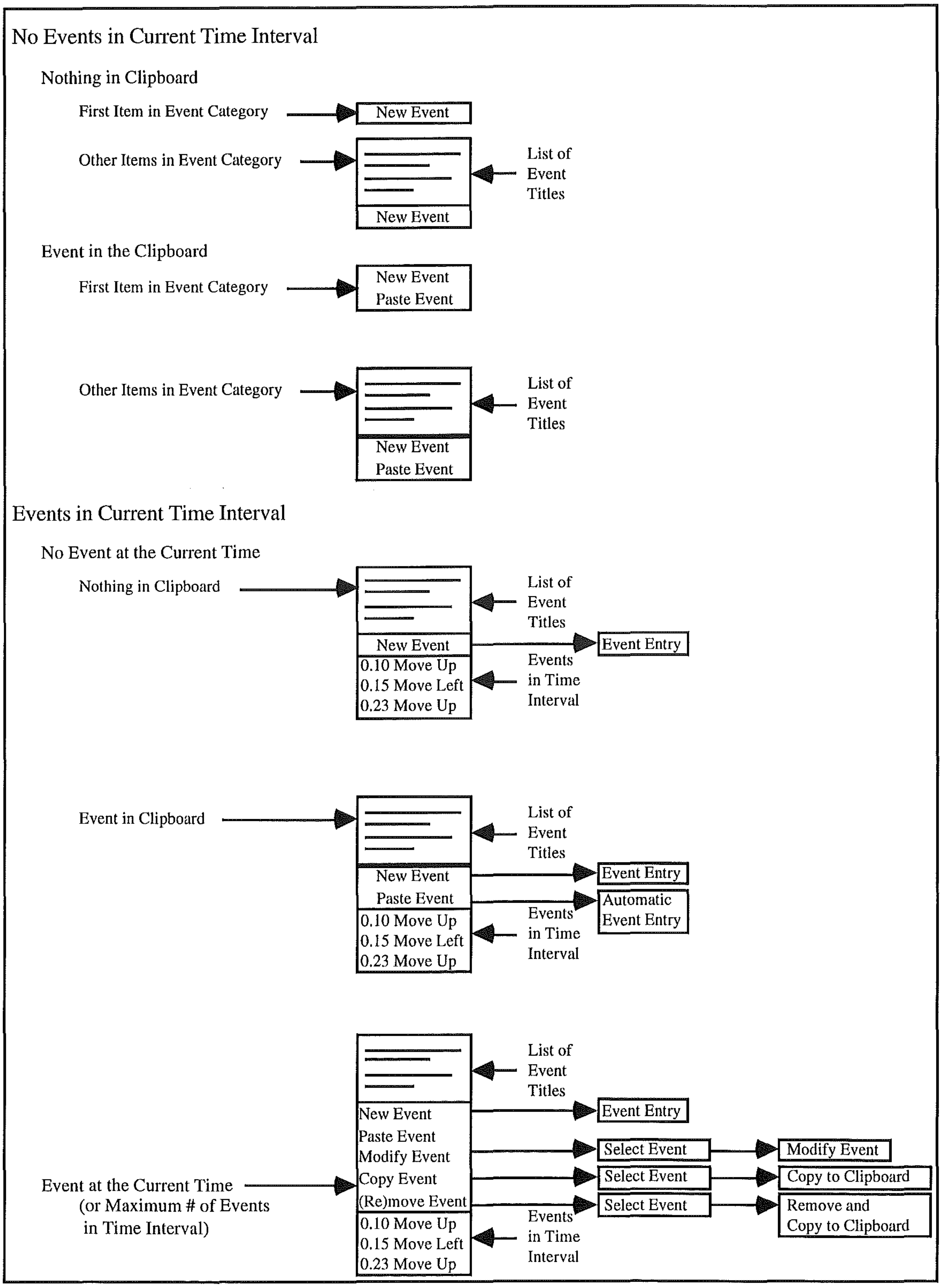
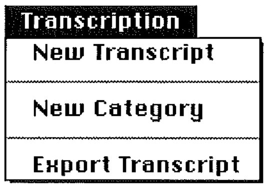


Figure 4.49. Menu Options Available With "Event Conflicts" Option Turned On.

Menus

The STANDARD mode of *Video Transcriptor* offers more enhanced capabilities with the *New Transcript* menu item.

Transcription Menu



New Transcript

Under the BASIC mode of *Video Transcriptor* when the user selected *New Transcript* a single transcription window would be opened. This single transcription window was used for all notes for a movie. The STANDARD mode offers the same global transcription window but also transcription windows for each track in the movie.

If an existing transcription window is available then it is activated. The menu command is context sensitive. It will create a new transcript window for the current track which is active. For example if a video window is the foremost window then a transcription window associated with this track is activated. If an audio track window is the foremost then a transcription window is activated for this track. There is one transcription window available for every track in the movie and one for the movie as a whole. This last transcription track is activated when the Master Controller is the foremost window and selecting the *New Transcript* command. There are two other ways to bring up a transcription window. These are; by clicking on the transcript button in the lower left corner of the video and audio windows, or by clicking the transcription window button in the Track Component on the Master Controller for the appropriate track.

Section 5

Expert Mode Features

Introduction

This Section shows you how to use the additional features that are available with *Video Transcriptor* in the EXPERT mode. Once again this section is supposed to be a self guiding tutorial of the enhanced features available. Section 3 - *Tutorial* discussed the features made available in the BASIC mode and Section 4 - *Standard Mode Features* discussed the features made available in the STANDARD mode.

Topics covered in this section:

- Variable Speed Controller
- Features to be Developed for Future Releases
    - Audio Window Features
    - Audio Component on the Master Controller
    - Tools for the Individual Track Windows
    - Speech Track Capabilities

This section is designed to complement Section 3 - *Tutorial* and Section 4 - *Standard Mode Features*. It assumes that the user understands all of the features available in the previous sections. Section 6 - *Expert Mode Features* introduces the ability to play back the movie data at any speed and direction. Other more complex features that are under development are discussed with a preview of what to expect in future releases of *Video Transcriptor*.

We shall start this section by discussing the variable speed controller in detail, figure 5.1, and then discuss the other features to be developed.

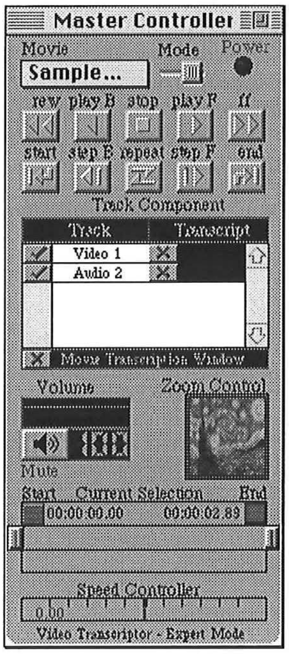


Figure 5.1. The Master Controller in EXPERT Mode.

Variable Speed Controller

Normal playback of a movie or stepping through a movie may be inappropriate or unacceptable when trying to examine the contents of a movie. The EXPERT mode of *Video Transcriptor* offers a control to allow variable speed playback control. This control is located at the bottom of the Master Controller, figure 5.2.

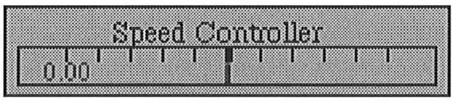


Figure 5.2. Variable Speed Controller.

The playback speed is variable from -5, reverse direction, to 5 times the normal speed. When the movie is stationary the control looks like figure 5.2. By dragging the green bar, located in the centre of the control, you control the speed and direction of the playback. Try dragging the bar to the right, figure 5.3.

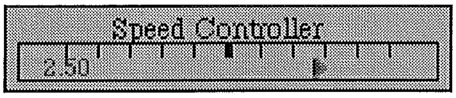


Figure 5.3. Increasing the Playback Speed.

As you drag the vertical bar to the right it changes shape into a directional triangle. This indicates the direction of playback. To the right of the centre mark the movie will playback in a normal direction, dragging it to the left will play the movie in reverse. As you drag the control slowly to the right you will notice three things happening:

- A speed indicator tells you what speed you are playing back the movie relative to a normal playback speed of 1.0. Initially it is 0 for no movement. As you drag it slowly to the right it will go through 0.5, half the normal playback speed, 1.0 which is normal playback speed, 2.0 which is two times the normal playback speed and so on. If you move the speed indicator to the left of the centre mark the indicator will be a negative number to represent the speed in reverse.

- The video windows will start to playback the video information at the appropriate playback speed and direction.

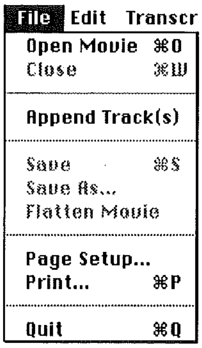
- The audio information is played back at the appropriate speed and direction that the movie is currently moving.

The variable speed controller allows the user to have total control of playback of the movie information in any direction and at any speed desired.

Menus

The EXPERT mode of *Video Transcriptor* adds the ability to add new tracks to a movie through the *Append Track(s)* menu item.

File Menu



Append Track(s)

After opening a movie file for the first time it will typically contain one video and one audio track. There may be times when the user wishes to analyse more than one subject. This would involve having more than one video and/or audio track, depending on the source material to be compared. This menu option enables the user to import individual tracks from a second movie. The creation of additional movies is done outside *Video Transcriptor* as already mentioned. First of all the second movie must be located. This is achieved by the standard Macintosh open dialog box as we saw in the *Open Movie* menu section. When the second movie is opened, a second dialog box is opened. This dialog box contains a list of all available tracks in the second movie. Figure 5.4 is an example of the list of available tracks that can be appended from the new movie file.

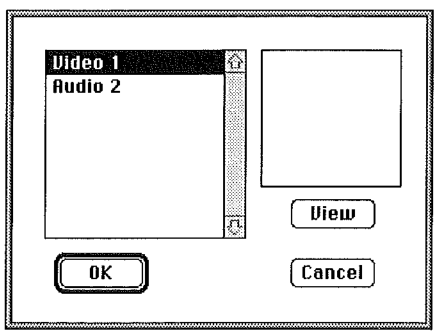


Figure 5.4. Dialog box showing list of tracks in a movie.

After selecting an individual track in the list, by pressing the *view* button, a sample of the individual track is played. If it is a video track, the contents will be shown in the box to the right of the list of tracks. If it is an audio track then it will be played through the computers speaker. By pressing the same button the sample will stop. If a track is found which is to be inserted, then pressing the *OK* button, when the track is selected, will insert the track into the current movie which is being analysed. The list of tracks is then updated to reflect this addition of a single track into the movie. If the *Cancel* button was pressed then the operation of inserting a track into the movie is aborted.

Features Under Development For Future Releases

Several features are currently under development and will be available for future release. These include:

- Audio representation in the audio windows.
- Audio component on the Master Controller.
- A set of tools for each track window.
- A speech track.
- More advanced exporting facilities.
- The ability to offset new tracks that have been appended to the movie file.

Most of these enhancements are related to audio analysis of the audio tracks.

Audio Representation in the Audio Windows

Current development involves analysis of the audio information in order to provide more facilities for the user. At present there is only the ability to playback the audio information at any speed and direction. This however does not provide adequate visual feedback for the user. The audio windows available at present do little more than confirm that the audio tracks are currently active. These audio windows will provide more visual information to the user, consisting of optionally peak or average audio levels of each individual audio track. This feature will greatly assist in detecting when audio noise begins and stops. At present the user has to guess the precise times but with a visual aid this will hopefully become a trivial task.

Audio Component on the Master Controller

Together with the advancements on the audio windows, more control of the audio tracks is needed on the Master Controller. A master volume control is available on the Master Controller below the Track Component. This consists of a mute button together with a volume slider control. The only control for individual audio tracks is using the Track Component to turn off the individual tracks. There is no control for adjusting the individual volume for each audio track. Each audio track will not contain audio information that is of a comparable volume to the other tracks. Each track needs to be adjusted to have a comparable volume. In other situations a user may wish to emphasise one particular track in relation to the others to make it clearer. These controls will become an integral part of the Track Component in a future release of *Video Transcriptor.*

Tools For Each Track Window

A desirable feature for the users was the capability of analysing multiple subject's simultaneously. This capability has been incorporated into *Video Transcriptor* already. This allows multiple subject's to be analysed or the same subject relative to other scenarios. One feature that is to be addressed is the ability to analyse different portions of separate tracks in the same movie. For example there may be two video tracks in the current movie being analysed. We can stop the movie at a specific location of interest. At this point we may wish to compare this current location of the video in one track with a segment of video that occurred thirty seconds ago in the second track. What we would like to be able to do is to reposition this second video track back thirty seconds to make a comparison. Another possibility is for the user wanting to make a comparison of a piece of video with a segment that occurred in the same video track earlier on.

Other features addressed as additional tools is an enhancement to the standard zoom controller. The standard zoom controller only addresses the movie as a whole, subsequently all video tracks are affected. An additional feature needed is zoom facilities to address individual video tracks independent of the others. These features are being addressed for future releases of *Video Transcriptor.*

Speech Track

This has two aspects to it, firstly to assist in the annotation of speech, and secondly the ability for the user to make speech annotations.

The transcription window is designed for taking detailed notes about specific events. The short description about each event that has occurred is displayed in the transcription window. A facility that may prove to be useful is the ability to transcribe any speech that may be on the audio tracks. The playback of the audio tracks would be able to provide the written version of the audio track appearing in real time. This is useful when trying to analyse the audio track in detail. The audio information itself may not be available due to the track being stopped. The audio display that has been proposed here may prove useful to see where the audio begins but there is still the problem of knowing what the actual sound is. By writing down what is being said the written words can be compared against when the actual sound begins and ends.

Another useful feature that could also apply to the other transcription notes is the ability to make speech annotations. This involves speaking into the computer's microphone about events instead of always having to type in all of the information. This provides alternative methods for the annotation of time-based information.

Once again these are methods that will be available in future releases of *Video Transcriptor.*

Exporting the Transcription Information

Exporting the results from the transcription process is necessary before the analysis stage of the results can commence. At present a text based file is produced with all of the relevant information available in a readable format. The next stage is to create the facilities for the user to define the exact format that the exported file should be in, the fields and the order in which they are written. The exported results can then be easily imported into statistical analysis packages, databases and other programs. As mentioned previously, *Video Transcriptor* is not designed to analyse the results of the notes which are taken but is designed to provide a simple means for which the results can be exported for use in more specialised packages.

Track Offsets

Another feature that is needed in the EXPERT mode of *Video Transcriptor* is the ability to offset individual tracks from each other. When a user uses the *Append Track(s)* menu item in the File Menu the track is automatically inserted so that it begins at time zero. The main reason for inserting additional tracks is the ability to then make comparisons between them. It is not guaranteed that the newly inserted track is going to be in synch with the existing tracks so there is a need to adjust the new track(s) from the beginning of the movie. For example the new track(s) may need to start two seconds after the movie actually begins just so the information is all aligned with each other.

Glossary

Acquisition: The process of transferring data from the analog to the digital world. This term is more specifically used to denote the process of digitisation of analog source video material.

Aliasing: The result of sampling a signal at less than twice its natural frequency. Aliasing causes data to be lost in the conversion that occurs when resampling an existing signal to form a new signal. Aliasing can be avoided, by sampling a signal at more than twice its natural frequency, a process referred to as "anti-aliasing" (see below). When aliasing occurs for digital images, there can be jagged line artifacts and missing detail; when aliasing occurs for sound, there can be a pronounced ragged "hiss" sound and subtle details in the sound will be lost. Aliasing is to be avoided as much as possible, by using proper filtering processes when resampling. Known aliasing situations to avoid, are sampling 22khz sound down to 11khz by dropping every other sample, and sampling 320 x 240 resolution video data down to 160 x 120 resolution by dropping every other pixel and every other scanline. Typically, aliasing occurs when a real-time process of resampling does not have enough time (or money) to spend properly sampling a signal to avoid the artifact.

Analog: A continuously, time varying signal. Quantities that change over time in the natural world are usually considered analog signals.

Anti-aliasing: The process of sampling a signal at more than twice its natural frequency, to ensure that aliasing artifacts (see above) do not occur. The anti-aliasing technique can be applied in real-time or non real-time. Anti-aliasing methods will improve the picture quality of stills or video, and the audio quality of digital sound or music. Anti-aliasing is sometimes also referred to as over sampling. Over sampling is used in some commercial CD audio players, where it is used to output a single audio sample as a combination of all of the neighbouring samples, each multiplied by a weighting coefficient.

Application: Short for application program. A program that performs a specific task, such as word processing, database management, or graphics.

Aspect ratio: The ratio of the width of an image to its height.

Audio: A recording of sound. In the context of *Video Transcriptor*, sound that can be recorded synchronised with the video from the video source. *Video Transcriptor* uses the facilities of the Macintosh to play the sound.

Audio stream: A continuous stream of audio data. An audio stream is generally considered a real-time object. The stream can come from a CD-ROM, hard disk drive, RAM or the network. It may be expected to run at a constant rate (or be buffered to appear as constant rate).

Audio/video

synchronisation: The process of maintaining the synchronisation between an audio track and a video track. The process ensures that audio and video are linked properly to permit showing the right frame of video at the same time that the sound corresponding to that frame is played. This is done independently of the underlying rate of the audio, and the underlying rate of the video. This will maintain lip sync for talking heads, and allow an image of an event to be in sync with the sound of the event.

Blockiness: An image compression artifact for various compression schemes, such as MPEG and the Apple Video Compressor. For methods that use frame differencing, blockiness appears as blocks of pixels that contain image data from previous frames at the corresponding locations in the picture. For methods that use motion estimation, blockiness appears as blocks of pixels that contain image data from previous frames at the differing locations in the picture. As temporal compression increases, the blockiness artifact worsens. If temporal compression is increased highly, then the artifact of blockiness will be prevalent ubiquitously.

Blurriness: An image compression artifact for various compression schemes, such as JPEG. As the compression ratio is increased, the image will tend to blur, or defocus, and important details in an image will appear smudged or illegible. The blurring is the result of inadequate colour or luminance resolution in the areas of the image.

Brightness: Brightness controls will affect the overall luminosity, or radiance of light energy from an image.

Capture: The process of transferring data from the analog to the digital world. This term is more specifically used to denoted the process of digitisation of analog source video material. Capture can either be in real-time ("live") or in non real-time ("controlled").

Clipboard: An area in the computer's memory that functions as a holding place for what was last cut or copied. Information if the Clipboard can be pasted into documents.

Codec: A compression/decompression system extension used by QuickTime to save disk space. Different codecs use different image compression schemes.

Component: A software entity, managed by the Component Manager in QuickTime, that provides a defined set of services to its clients. Examples include clock components, movie controller components, and image compressor components.

Compressor component: General term used to refer to both image compressor components and image decompressor components.

Contrast: The ratio between the lightest and darkest intensities in an image. The higher the contrast, the better the definition the detail will be.

Controlled capture: The process of capturing video from the analog world, using a computer to control a controllable video source. The source might be a frame accurate videotape recorder, or a laser disk player. Controlled capture is typically in a non real-time process.

Decimation: The process of creating a lower resolution version of a signal. Frequently used to describe the creation of a lower resolution version of an image, such as reducing a quarter screen image into a sixteenth screen image. Decimation can be synonymous with creating a lower resolution version of an image without sampling at twice the natural frequency of the image in the process. This type of decimation will result in aliasing, and produce objectionable image artifacts, particularly ragged edges and image jitter.

Deinterleaving: The process of reading a stream of interleaved audio/video, and parcelling out the video for display on a monitor, and the audio for output to a speaker. Deinterleaving requires the data stream to be broken up into smaller pieces of individual media, and sent out to the appropriate media handlers for display.

Delta frame: This is a term used to describe a particular kind of frame used with frame differencing. The delta frame is a frame that contains only differences between the current frame it represents, and the frame that precedes it. A delta frame tends to be considerably smaller than an intraframe (see below). A stream of temporally compressed data must be interspersed with key frames (also called intrafames, or stand-alone frames) in addition to delta frames, so that audio/video synchronisation can occur.

Dialog Box: A box that contains a message, often requesting more information from the user. Sometimes the message is accompanied by a beep.

Difference frame: This term is synonymous with delta frame.

Digital multimedia: A generic term used to describe the integration of still images, audio, video, text, graphics, animation in a single presentation or file, with all the aforementioned media in digital format. Frequently, the digital media must be stored in compressed form, to facilitate compact storage and real-time display. An alternate term is media integration.

Dither: A technique used to display an image at N bits per pixel display, that exists at a higher bit depth representation. A common usage is to dither a 24 bit per pixel image for display on an 8 bit screen. The method is used to improve the picture quality at the lower bit depth, using techniques of either ordered dither (see below) or error diffusion dither.

The technique of making adjacent pixels different colours to give the illusion of a third colour. Dithering can produce the effect of a shades of grey on a black-and-white display, or more colours on an 8-bit (256 colours) colour display.

Double-click: Two clicks in quick succession, interpreted as a single command. The action of a double-click is different from that of a single-click. For example, clicking an icon selects the icon; double-clicking an icon opens it.

Duration: The playing time of a clip, measured in the standard format approved by the Society of Motion Picture and Television Engineers (SMPTE format): Hours:Minutes:Seconds.Frames.

Dynamic Data: The data type for QuickTime movies. It corresponds to the text and graphics data types previously defined for the Macintosh. All three data types are transparent, meaning you can cut and paste them in any application.

Filtering: The process of modifying an input signal to produce an output signal. Filtering can be applied to images, sounds, etc. It is commonly used to maintain signal information, when higher or lower resolution representations are being generated. High quality filtering uses much information from neighbouring sample data to produce an output signal. Low quality filtering uses very little information from neighbouring samples.

Flattening: The process of copying all of the original data that is referred to by references in QuickTime tracks into a QuickTime movie file. This can also be called resolving references. Flattening is used to bring in all of the data that may be referred to from multiple files, after QuickTime editing is complete. It makes a QuickTime movie "whole", or standalone, in that it can be played on any system, without requiring any additional QuickTime movie files or tracks, even if the original file referenced hundreds of files. The flattening operation is essential to be performed before mastering a QuickTime movie for CD-ROM.

Frame: The basic unit of information in television, video, and QuickTime movies; a single image. The specifications for frames vary according to the television format that is used. A single image in a sequence of images.

Frame differencing: A form of temporal compression that involves examining redundancies between adjacent frames in a moving image sequence. Frame differencing can improve compression ratios considerably for a video sequence.

Frame dropping: The process of dropping video frames to stay synchronised with a soundtrack. Ideally, when a sequence is played, not many video frames will be dropped. Frame dropping will increase when the hard disk it is being played from is fragmented, or if there are many events occurring on a user system, other than the playback of a movie. Synonym for frame skipping.

Frame rate: The rate at which a movie is displayed. This term shows the number of frames per second that are actually being displayed.

In QuickTime, the frame rate at which a movie was recorded, may be different than the frame rate at which it is being displayed. On very fast machines, the playback frame rate may be faster than the record frame rate. On slow machines, the playback frame rate may be slower than the record frame rate. Frame rates may be fractional.

Frame skipping: The process of dropping video frames to stay synchronised with a soundtrack. Ideally, when a sequence is played, not many video frames will be dropped. Frame dropping will increase when the hard disk it is being played from is fragmented, or if there are many events occurring on a user system, other than the playback of a movie. synonym for frame dropping.

Hardware compression: Using dedicated VLSI, RISC or DSP to compress or decompress video/audio streams. hardware compression can come in two forms - fixed function or programmable, and vendors are exploring both approaches. Generally, hardware compression should be used.

Hue: Similar to "tint control" on a television set. Rotates the colours.

Image compression manager: A component that provides image compression services.

Image decompression manager: A component that provides image decompression services.

Image sequence: A series of visual representations usually represented by video over time. Image sequences may be also generated synthetically, such as from an animation sequence.

Interframe coding: Similar to frame differencing, but may use different techniques to exploit commonality between frames. For example, the

MPEG standard uses a technique called motion estimation to find common regions between adjacent frames. These common regions may not overlap in the spatial domain, and the process of decompression involves shifting the common blocks to the correct position.

Interlacing: A video mode that updates half the scan lines at a time and goes through the second half during the next pass.

Interleaving: The process of combining streams of different types (such as video together with audio, or MIDI together with SMPTE, or video with SMPTE) together into a storage format. The data is interleaved in time, so that N seconds of video may follow M seconds of audio. Sometimes interleaved data streams have regular patterns with a certain periodicity, and sometimes the data streams will have irregular patterns.

A technique for larger movies in which sound and video data may be alternated in small pieces, so the data can be read off disk as it is needed. Interleaving allows for movies of almost any length with little delay on start-up.

Interpolation: The process of extrapolating a signal to create more data than exists in the original signal. Interpolation may be used to resize an image to be larger than its current size; interpolation may also be used to fill in missing audio samples, if some samples are missing. There are various quality levels that can be used for the interpolation process. Like good filtering methods, good interpolation methods will use data from many neighbouring samples to perform the interpolation operation.

Intraframe coding: Intraframe coding compresses only a single frame. The process does not require looking at adjacent frames in time to achieve compression. Intraframe coding allows fast random access and reverse play. The intraframes themselves are sometimes also called "standalone" frames, since they can be interpreted and displayed without knowledge of neighbouring frames.

JPEG: Stands for Joint Photographic Experts Group. JPEG is the recently codified International Standards Organisation (ISO) standard for compressed still images. This standard permits images to be compressed as to be visually lossless at 10:1 compression ratios from a 24-bit source image. The version of JPEG supplied with QuickTime complies with the baseline ISO standard bit stream, version 9R9. This algorithm is best suited for use with natural images.

Key frame: A sample in a sequence of temporally compressed samples that does not rely on other samples in the sequence for any of its information. Key frames are placed into temporally compressed sequences at a frequency that is determined by the key frame rate. Typically, the term key frame is used with respect to temporally compressed sequences of image data. See also sync sample.

Key frame rate: The rate at which key frames are generated for a stream of temporally compressed video. Keyframes are standalone frames that are used for fast random access and reverse play. They must be used in a stream of temporally compressed video to maintain audio/video synchronisation.

Letterbox: The format used to display wide-aspect ratio film images on NTSC videotapes and laser disks. The letterbox format causes black borders to appear at the top and bottom of the NTSC image, with the film content filling the middle region of the screen in a roughly 2:1 aspect ratio format.

Live capture: The process of capturing analog video to digital form in real-time. Also used to describe the process of capturing both analog video and audio. Live capture can be accomplished by storing data into RAM or hard disk. Generally speaking, live capture at good quality requires either hardware compression, a fast CISC machine or a RISC processor.

Lossless compression: A type of data compression scheme that preserves all of the original data. Image-compression schemes attempt to recreate source image (or sound) data with varying degrees of fidelity.

Lossy compression: A type of image-compression data that does not preserve the data precisely; image data is lost, and it cannot be recovered after compression. Most lossy schemes try to compress the data as much as possible, without decreasing the image quality in a noticeable way.

Lossless recompression: The process of compressing an already compressed image with no degradation in picture quality. This permits editing and image processing of compressed still image or video data with no loss in quality, allowing multiple generations of manipulation of the image to occur. This type of feature is very useful for post-production and special effects, as well as the need to change compression parameters after a sequence has already been compressed.

Image sequence: A sequence of images over time. Video is an alternate term, but an image sequence may also be generated synthetically, such as from an animation sequence.

Media: A Movie Toolbox data structure that contains information that describes the data for a track in a movie. Note that a media does not contain its data; rather, a media contains a reference to its data, which may be stored on disk, CD-ROM disc, or any other mass storage device.

MooV: The resource name for QuickTime movie resources. Files of type Moov can be read and written by standard QuickTime applications, for editing, playback, processing, etc.

Motion estimation: The process of searching a fixed region of a previous frame to find a matching block of pixels with a block of pixels of the same size under consideration in the current frame. The process involves an exhaustive search for many blocks surrounding the current block from the previous frame. Motion estimation is highly compute intensive, and can require as much as one billion operations per second to process real-time video, making the technique ill-suited to today's microprocessors for real-time capture. Motion estimation is used to achieve very high compression ratios, since it stores a motion vector for each block, instead of any other pixel or transform data, and the method is used in the world-wide standard for compressed video, MPEG.

Motion compensation: The process of playing back motion estimated data when decompressing video. Motion compensation is the act of examining the motion vectors generated by motion estimation, and moving the pixel data from the previous frame from the correct position to the current position on the screen. Motion compensation is a great deal less computationally intensive than motion estimation.

Movie: Used here to describe a QuickTime movie (may also be used to describe a movie viewed at a movie theatre). A QuickTime movie adheres to a well-defined hierarchical data structure of movie, tracks and media. The Movie may reside in RAM or on disk. A QuickTime movie can contain any number of tracks of different types of temporal data. The QuickTime movie format is a published specification for the storage of temporal data.

Movie file: The disk based instantiation of a QuickTime movie. A movie file can be either single-fork or double-fork, and can be moved across different computing platforms. Typically movie files contain many different versions of the same basic material (such as edits), without requiring need for replicating that material. In this case, the movie files all point to the same data.

MPEG: A recently codified standard for the compression and decompression of full-motion video. The standard was developed by the ISO based Motion Pictures Experts Group.

NTSC: *(National Television System Committee).* A broadly used standard used for representing analog video. The standard was developed in the 1940's, essentially unchanged today. This colour-encoding method was adopted by the NTSC in 1953. This was the first monochrome-compatible, simultaneous colour transmission system used for public broadcasting. NTSC is in use in both the US and Japan. NTSC resolution contains 525 lines and runs at a rate of 59.96 fields per second, or 29.97 frames per second. Nevertheless, we generally refer to NTSC video running at 30 fps. NTSC is typically generated at 640 x 480 resolution. NTSC video is interlaced, meaning that alternate fields each contain alternating scanlines of data to form a full frame.

Nyquist frequency: A sampling rate that is equal to twice the highest frequency contained in all of the samples in a signal. Anti-aliasing requires sampling at the Nyquist frequency or higher.

Ordered dither: A method used for dithering, that uses a fixed pattern. Ordered dither will improve the quality of the display of images, but my still have an artifact of a patterned look to it. Is less expensive to calculate than error diffusion dither, but produces lower quality.

p*64: A standard used for compressed video, targeted at applications for telecommunications. The standard can run at many different bit rates from 64 kbits/sec up to 1.5 mbits/sec, where p = 1,2, ...24. Requires substantial dedicated silicon and hardware.

PAL: *(Phase Alternating Line).* A European standard (Great Britain) used for representing analog video. A colour encoding system used widely in Europe in which one of the subcarrier phases derived from the colour burst is inverted in phase from one line to the next. This technique minimises hue errors which may result during colour video transmission. PAL resolution is 768 x 576 resolution and runs at 50 fields per second, or 25 frames per second. PAL video is interlaced, meaning that alternate fields each contain alternating scanlines of data to form a full frame.

PC-VCR: A video cassette recorder and playback device manufactured by NEC, that can be controlled by a computer via the serial port, and has the capability the read and write time code with analog videotapes. The PC-VCR can be used to play sequences of video from one specific timecode to another; it can also be used capture QuickTime movies in controlled capture mode with good picture quality.

Pixel: *(Short for picture element).* A single dot on a computer display or in a digital image.

Pixellation: The appearance of an image that has large square blocks of constant colour. The larger the blocks are, the more pixellated the image appears.

Play all frames: A mode in QuickTime that plays a sequence of video frames directly from hard disk or CD-ROM without applying any synchronisation techniques. Generally speaking, "play all frames" mode will play much faster than the standard mode of using audio/video synchronisation.

Playback: The process of displaying images, video, animation, sound or other media together or individually in real-time from a storage media or network.

Post-processing: Used here to denote processing of video or sound data after a QuickTime movie has been captured. Post-processing can be performed as background task, or off-line. Synonymous in some regards with transcoding.

Quantization: The process of representing information with less bits than was used in the original representation. For example, using 16 bits per pixel for an image to represent an image generated at 24 bits per pixel is considered quantization. Quantization may result in image artifacts, such as banding or contouring.

QuickTime: Apple's new protocol for real-time multimedia. QuickTime integrates moving images and sound in a single, new data type called dynamic data. The new data type makes multimedia data as transparent as text and graphics, so that the data can be cut and pastes in all documents as easily as text and graphics.

Random-access-memory (RAM): The part of a computer's memory that stores information temporarily while you are working with it.

Raw: A synonym for uncompressed video data. Can be used with the Apple None compressor.

Recompression: The act of compressing image or video data that has already been compressed.

Resize: Enlarging or shrinking an image. If resizing is accompanied by proper filtering, then image quality at the different sizes will be good; if not filtering is used, image quality will be low.

Scalable video: A technique used for representing compressed video. Scalability can be used in two ways - constant frame rate with variable image quality, or variable frame rate with constant image quality. The usage of scalable implies that the video stream can be played back on a variety of differing computational platforms, and still produce similar results. Scalability may also be associated with bandwidth, so that the same video stream can work over different telecommunications lines, independent of the available bandwidth. Much research is in progress on scalable video.

SECAM: *(Systeme Electronique Couleur avec Memoire)* - Sequential Colour With Memory. A European standard (France and Eastern Europe) used for representing analog video. SECAM resolution is 768 x 576 resolution and runs at 50 fields per second, or 25 frames per second. SECAM video is interlaced, meaning that alternate fields each contain alternating scanlines of data to form a full frame. SECAM uses a colour encoding scheme in which the red and blue colour-difference information is transmitted on alternate lines, requiring a one-line memory in order to decode green information.

Spatial compression: Image compression that is performed with in the context of a single frame. This compression technique takes advantage of redundancy in the image to reduce the amount of data required to accurately represent the image.

Software compression: The compression of images or video sequences using a general-purpose microprocessor. This approach has the advantage of being available to all users, is flexible, programmable and will only get faster over time, as computers get faster over time. Hardware compression is typically used to get order of magnitude or more improvements in performance beyond the capabilities of hardware compression.

Standalone frame: Synonym for intraframe. A video frame that can be decompressed without examining any neighbouring video frames. Used for random access, reverse play and synchronisation.

Streaking: An artifact generated by using the Apple Animation Compressor in lossy mode. Tends to generate long lines of constant colour across scanlines, which may be of incorrect colour. The more lossy the quality levels are set to with this compressor, the more pronounced the streaking will be. Also - strange trend in the early 1970s.

Streaming from disk: The process of reading data directly from a hard disk or CD-ROM, without pausing or searching on the disk.

S-video: A video-format in which colour and brightness information are encoded as separate signals. The s-video format is known as component video as opposed to composite video, which is the NTSC and PAL format.

Temporal compression: Image compression that is performed between frames in a sequence. This compression technique takes advantage of redundancy between adjacent frames in a sequence to reduce the amount of data that is required to accurately represent each frame in the sequence. Sequences that have been temporally compressed typically contain key frames at regular intervals. See also spatial compression.

Time base: A set of values that define the time basis for an entity, such as a QuickTime movie. A time base consists of a time co-ordinate system (that is, a time scale and a duration) along with a rate value. The rate value specifies the speed with which time passes for the time base.

Time co-ordinate system: A set of values that defines the context for a time base. A time co-ordinate system consists of a time scale and a duration. Together, these values define the co-ordinate system in which a time value or a time base has meaning.

Time scale: The number of time units that pass per second in a time co-ordinate system. A time co-ordinate system that measures time in sixtieths of a second, for example, has a time scale of 60.

Time unit: The basic unit of measure for time in a time co-ordinate system. The value of the time unit for a time co-ordinate system is represented by the formula (1/time scale) seconds. A time co-ordinate system that has a time scale of 60 measures time in terms of sixtieths of a second.

Time value: A value that specifies a number of time units in a time co-ordinate system. A time value may contain information about a point in time or about a duration.

Track: A Movie Toolbox data structure that represents a single data stream in a QuickTime movie. A movie may contain one or more tracks. Each track is independent of other tracks in the movie and represents its own data stream. Each track has a corresponding media. The media describes the data for the track.

Transcoding: The process of converting a still image or video sequence from one format to another. There are many operations that can be applied when transcoding, including modification of picture size, bit depth, compression parameters and more. However, transcoding is typically referred to as a means of converting video sequences from one compression format (or set of parameters) to another compression format (or set of parameters).

Variable bit rate coder: A compression scheme that generates variable sizes for different frames of video, or different samples of audio. Variable bit rate coders are common for both spatial and temporal compression.

Video: A recording of a visual image that can be shown on cathode ray tubes of a kind in computer displays, television sets, and video monitors.

Video stream: A continuous stream of video data. May be stored on disk, or be retrieved from a network or CD-ROM.

Virtual memory: The memory space that is separate from the main memory (physical random-access-memory), such as hard disk space. Virtual memory allows you to work on large documents without requiring you to have large amounts of RAM.

Acknowledgements

I would like to acknowledge:

- Dr. Timothy Bell — my supervisor,
- Pari Kolahi — Psychology Department research student,
- Marcus Finlay — tester,
- Fellow masters students — ideas and testing.

The author would also like to thank Apple Computer for the development of QuickTime which makes *Video Transcriptor* a reality.

Trademarks

Macintosh and QuickTime are trademarks of Apple Computer, Inc.